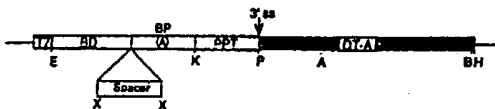| (51) International Patent Classification [7] : C12N 15/90, A61K 48/00, C12N 15/11 // A61P 11/00	A2	(11) International Publication Number: **WO 00/09734**
		(43) International Publication Date: 24 February 2000 (24.02.00)

(54) Title: METHODS AND COMPOSITIONS FOR USE IN SPLICEOSOME MEDIATED RNA TRANS–SPLICING

(57) Abstract

The molecules and methods of the present invention provide a means for *in vivo* production of a trans–spliced molecule in a selected subset of cells. The pre–trans–splicing molecules of the invention are substrates for a trans–splicing reaction between the pre–trans–splicing molecules and a pre–mRNA which is uniquely expressed in the specific target cells. The *in vivo* trans–splicing reaction provides a novel mRNA which is functional as mRNA or encodes a protein to be expressed in the target cells. The expression product of the mRNA is a protein of therapeutic value to the cell or host organism a toxin which causes killing of the specific cells or a novel protein not normally present in such cells. The invention further provides PTMs that have been genetically engineered for the identification of exon/intron boundaries of pre–mRNA molecules using an exon tagging method. The PTMs of the invention can also be designed to result in the production of chimeric RNA encoding for peptide affinity purification tags which can be used to purify and identify proteins expressed in a specific cell type.

METHODS AND COMPOSITIONS FOR USE IN SPLICEOSOME
MEDIATED RNA TRANS-SPLICING

SPECIFICATION

1. INTRODUCTION

5   The present invention provides methods and compositions for

generating novel nucleic acid molecules through targeted spliceosomal *trans*-splicing.

The compositions of the invention include pre-*trans*-splicing molecules (PTMs)

designed to interact with a natural target precursor messenger RNA molecule (target

pre-mRNA) and mediate a *trans*-splicing reaction resulting in the generation of a

10   novel chimeric RNA molecule (chimeric RNA). The PTMs of the invention are

genetically engineered so as to result in the production of a novel chimeric RNA

which may itself perform a function, such as inhibiting the translation of the RNA, or

that encodes a protein that complements a defective or inactive protein in a cell, or

encodes a toxin which kills specific cells. Generally, the target pre-mRNA is chosen

15   as a target because it is expressed within a specific cell type thus providing a means

for targeting expression of the novel chimeric RNA to a selected cell type. The

invention further relates to PTMs that have been genetically engineered for the

identification of exon/intron boundaries of pre-mRNA molecules using an exon

tagging method. In addition, PTMs can be designed to result in the production of

20   chimeric RNA encoding for peptide affinity purification tags which can be used to

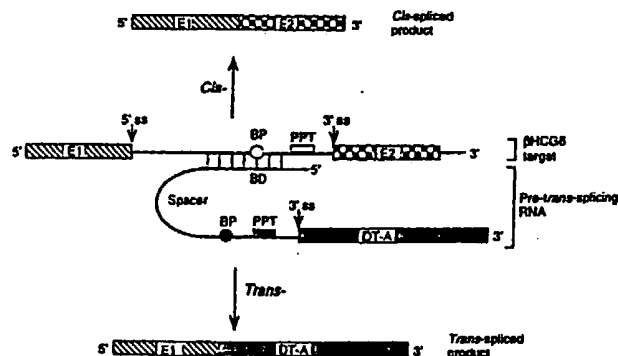purify and identify proteins expressed in a specific cell type. The methods of the

invention encompass contacting the PTMs of the invention with a target pre-mRNA

under conditions in which a portion of the PTM is *trans*-spliced to a portion of the

target pre-mRNA to form a novel chimeric RNA molecule. The methods and

25   compositions of the invention can be used in cellular gene regulation, gene repair and

suicide gene therapy for treatment of proliferative disorders such as cancer or

treatment of genetic, autoimmune or infectious diseases. The methods and

compositions of the invention can also be used to map intron-exon boundaries and to identify novel proteins expressed in any given cell.

2. BACKGROUND OF THE INVENTION

DNA sequences in the chromosome are transcribed into pre-mRNAs which contain coding regions (exons) and generally also contain intervening non-coding regions (introns). Introns are removed from pre-mRNAs in a precise process called splicing (Chow *et al.*, 1977, Cell 12:1-8; and Berget, S.M. *et al.*, 1977, Proc. Natl. Acad. Sci. USA 74:3171-3175). Splicing takes place as a coordinated interaction of several small nuclear ribonucleoprotein particles (snRNP's) and many protein factors that assemble to form an enzymatic complex known as the spliceosome (Moore *et al.*,1993, in The RNA World, R.F. Gestland and J.F. Atkins eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.); Kramer, 1996, Annu. Rev. Biochem., 65:367-404; Staley and Guthrie, 1998, Cell 92:315-326).

Pre-mRNA splicing proceeds by a two-step mechanism. In the first step, the 5' splice site is cleaved, resulting in a "free" 5' exon and a lariat intermediate (Moore, M.J. and P.A. Sharp, 1993, Nature 365:364-368). In the second step, the 5' exon is ligated to the 3' exon with release of the intron as the lariat product. These steps are catalyzed in a complex of small nuclear ribonucleoproteins and proteins called the spliceosome.

The splicing reaction sites are defined by consensus sequences around the 5' and 3' splice sites. The 5' splice site consensus sequence is AG/GURAGU (where A=adenosine, U = uracil, G = guanine, C = cytosine, R = purine and / = the splice site). The 3' splice region consists of three separate sequence elements: the branch point or branch site, a polypyrimidine tract and the 3' splice consensus sequence (YAG). These elements loosely define a 3' splice region, which may encompass 100 nucleotides of the intron upstream of the 3' splice site. The branch point consensus sequence in mammals is YNYUR_AC (where N = any nucleotide, Y= pyrimidine). The underlined A is the site of branch formation (the BPA = branch point adenosine). The 3' splice consensus sequence is YAG/G. Between the branch point and the splice site there is usually found a polypyrimidine tract, which is

important in mammalian systems for efficient branch point utilization and 3' splice site recognition (Roscigno, R., F. *et al.*, 1993, J. Biol. Chem. 268:11222-11229). The first YAG trinucleotide downstream from the branch point and polypyrimidine tract is the most commonly used 3' splice site (Smith, C.W. *et al.*, 1989, Nature 342:243-247).

In most cases, the splicing reaction occurs within the same pre-mRNA molecule, which is termed *cis*-splicing. Splicing between two independently transcribed pre-mRNAs is termed *trans*-splicing). *Trans*-splicing was first discovered in trypanosomes (Sutton & Boothroyd, 1986, Cell 47:527; Murphy *et al.*, 1986, Cell 47:517) and subsequently in nematodes (Krause & Hirsh, 1987, Cell 49:753); flatworms (Rajkovic *et al.*, 1990, Proc. Nat'l. Acad. Sci. USA, 87:8879; Davis *et al.*, 1995, J. Biol. Chem. 270:21813) and in plant mitochondria (Malek *et al.*, 1997, Proc. Nat'l. Acad. Sci. USA 94:553). In the parasite *Trypanosoma brucei*, all mRNAs acquire a splice leader (SL) RNA at their 5' termini by *trans*-splicing. A 5' leader sequence is also *trans*-spliced onto some genes in *Caenorhabditis elegans*. This mechanism is appropriate for adding a single common sequence to many different transcripts.

The mechanism of *trans*-splicing, which is nearly identical to that of conventional *cis*-splicing, proceeds via two phosphoryl transfer reactions. The first causes the formation of a 2'-5' phosphodiester bond producing a 'Y' shaped branched intermediate, equivalent to the lariat intermediate in *cis*-splicing. The second reaction, exon ligation, proceeds as in conventional *cis*-splicing. In addition, sequences at the 3' splice site and some of the snRNPs which catalyze the *trans*-splicing reaction, closely resemble their counterparts involved in *cis*-splicing.

Trans-splicing may also refer to a different process, where an intron of one pre-mRNA interacts with an intron of a second pre-mRNA, enhancing the recombination of splice sites between two conventional pre-mRNAs. This type of *trans*-splicing was postulated to account for transcripts encoding a human immunoglobulin variable region sequence linked to the endogenous constant region in a transgenic mouse (Shimizu *et al.*,1989, Proc. Nat'l. Acad. Sci. USA 86:8020). In addition, *trans*-splicing of c-myb pre-RNA has been demonstrated (Vellard, M. et al.

4

Proc. Nat'l. Acad. Sci. 89:2511-2515) and more recently, RNA transcripts from cloned

SV40 *trans*-spliced to each other were detected in cultured cells and nuclear extracts

(Eul *et al.*, 1995, EMBO. J. 14:3226). However, naturally occurring *trans*-splicing of

mammalian pre-mRNAs is thought to be an exceedingly rare event. The reaction

5   mechanism of *trans*-splicing is believed to be nearly identical to conventional *cis*-

splicing. It proceeds via the formation of a 2'-5' phosphodiester bond producing a 'Y'

shaped branched intermediate (equivalent to the lariat intermediated in *cis*-splicing).

             In vitro trans-splicing has been used as a model system to examine the

mechanism of splicing by several groups (Konarska & Sharp, 1985, Cell 46:165-171

10   Solnick, 1985, Cell 42:157; Chiara & Reed, 1995, Nature 375:510; Pasman and

Garcia-Blanco, 1996, Nucleic Acids Res. 24:1638). Reasonably efficient *trans*-

splicing (30% of *cis*-spliced analog) was achieved between RNAs capable of base

pairing to each other, splicing of RNAs not tethered by base pairing was further

diminished by a factor of 10. Other *in vitro trans*-splicing reactions not requiring

15   obvious RNA-RNA interactions among the substrates were observed by Chiara &

Reed (1995, Nature 375:510), Bruzik J.P. & Maniatis, T. (1992, Nature 360:692) and

(Bruzik J.P. and Maniatis, T., 1995, Proc. Nat'l. Acad. Sci. USA 92:7056-7059).

These reactions occur at relatively low frequencies and require specialized elements,

such as a downstream 5' splice site or exonic splicing enhancers.

20   In addition to splicing mechanisms involving the binding of multiple

proteins to the precursor mRNA which then act to correctly cut and join RNA, a third

mechanism involves cutting and joining of the RNA by the intron itself, by what are

termed catalytic RNA molecules or ribozymes. The cleavage activity of ribozymes

has been targeted to specific RNAs by engineering a discrete "hybridization" region

25   into the ribozyme. Upon hybridization to the target RNA, the catylytic region of the

ribozyme cleaves the target. It has been suggested that such ribozyme activity would

be useful for the inactivation or cleavage of target RNA *in vivo*, such as for the

treatment of human diseases characterized by production of foreign of aberrant RNA.

The use of antisense RNA has also been proposed as an alternative mechanism for

30   targeting and destruction of specific RNAs. In such instances small RNA molecules

are designed to hybridize to the target RNA and by binding to the target RNA prevent

5

translation of the target RNA or cause destruction of the RNA through activation of
nucleases.

Until recently, the practical application of targeted *trans*-splicing to
modify specific target genes has been limited to group I ribozyme-based mechanisms.
5   Using the *Tetrahymena* group I ribozyme, targeted *trans*-splicing was demonstrated in
E. coli. coli (Sullengen B.A. and Cech. T.R., 1994, Nature 341:619-622) , in mouse
fibroblasts (Jones, J.T. et al., 1996, Nature Medicine 2:643-648), human fibroblasts
(Phylacton, L.A. et al. Nature Genetics 18:378-381) and human erythroid precursors
(Lan et al., 1998, Science 280:1593-1596). While many applications of targeted RNA
10   *trans*-splicing driven by modified group I ribozymes have been explored, targeted
trans-splicing mediated by native mammalian splicing machinery, *i.e.*, spliceosomes,
has not been previously reported.

3. SUMMARY OF THE INVENTION

The present invention relates to compositions and methods for
15   generating novel nucleic acid molecules through spliceosome-mediated targeted *trans*-
splicing. The compositions of the invention include pre-*trans*-splicing molecules
(hereinafter referred to as "PTMs") designed to interact with a natural target pre-
mRNA molecule (hereinafter referred to as "pre-mRNA") and mediate a spliceosomal
trans-splicing reaction resulting in the generation of a novel chimeric RNA molecule
20   (hereinafter referred to as "chimeric RNA"). The methods of the invention encompass
contacting the PTMs of the invention with a natural target pre-mRNA under
conditions in which a portion of the PTM is spliced to the natural pre-mRNA to form
a novel chimeric RNA . The PTMs of the invention are genetically engineered so that
the novel chimeric RNA resulting from the *trans*-splicing reaction may itself perform
25   a function such as inhibiting the translation of RNA, or alternatively, the chimeric
RNA may encode a protein that complements a defective or inactive protein in the
cell, or encodes a toxin which kills the specific cells. Generally, the target pre-mRNA
is chosen because it is expressed within a specific cell type thereby providing a means
for targeting expression of the novel chimeric RNA to a selected cell type. The target
30   cells may include, but are not limited to those infected with viral or other infectious

6

agents, benign or malignant neoplasms, or components of the immune system which are involved in autoimmune disease or tissue rejection. The PTMs of the invention can also be genetically engineered to tag exon sequences in a mRNA molecule as a method for identifying intron/exon boundaries in target pre-mRNA. The invention

5   further relates to the use of PTM molecules that are genetically engineered to encode a peptide affinity purification tag for use in the purification and identification of proteins expressed in a specific cell type. The methods and compositions of the invention can be used in gene regulation, gene repair and targeted cell death. Such methods and compositions can be used for the treatment of various diseases including,

10   but not limited to, genetic, infectious or autoimmune diseases and proliferative disorders such as cancer.

4. BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A. Model of Pre-*Trans*-splicing RNA.

Figure 1B. Model PTM constructs and targeted *trans*-splicing strategy.

15   Schematic representation of the first generation PTMs (PTM+Sp and PTM-Sp). BD, binding domain; NBD, non-binding domain; BP, branch point; PPT, pyrimidine tract; ss, splice site and DT-A, diphtheria toxin subunit A. Unique restriction sites within the PTMS are indicated by single letters: E. coli, EcoRI; X, Xhol; K, Kpnl; P, Pstl; A, Accl; B, BamHI and Heterologous, HindIII.

20   Figure 1C. Schematic drawing showing the binding of PTM+Sp via conventional Watson Crick base pairing to the βHCG6 target pre-mRNA and the proposed *cis*- and *trans*-splicing mechanism.

Figure 2A. *In vitro trans*-splicing efficiency of various PTM constructs into βHCG6 target. A targeted binding domain and active splice sites

25   correlate with PTM *trans*-splicing activity. Full length targeted (pcPTM+Sp), non-targeted (PTM-Sp) and the splice mutants [Py(-)AG(-) and BP(-)Py(-)AG(-)] PTM RNAs were added to splicing reactions containing βHCG6 target pre-mRNA. The products were RT-PCR amplified using primers βHCG-F (specific for target βHCG6 exon 1) and DT-5R (complementary to DT-A) and analyzed by electrophoresis in a

30   1.5% agarose gel.

Figure 2B. *In vitro trans*-splicing efficiency of various PTM constructs. Full length PTM with a spacer between the binding domain and splice site (PTM+Sp), PTM without the spacer region (PTM+) and short PTMs that contain a target binding domain (short PTM+) or a non-target binding region (PTM-) were

5 added to splicing reactions containing βHCG target pre-mRNA. The products were RT-PCR amplified using primers βHCG-F and DT-3. For reactions containing the short PTMs, the reverse PCR primer was DT-4, since the binding site for DT-3 was removed from the PTM.

Figure 3. Nucleotide sequence demonstrating the *in vitro trans*-spliced

10 product between a PTM and target pre-mRNA. The 466 bp *trans*-spliced RT-PCR product from Figure 2 (lane 2) was re-amplified using a 5' biotin labeled forward primer (βHCG-F) and a nested unlabeled reverse primer (DT-3R). Single stranded DNA was purified and sequenced directly using toxin specific DT-3R primer. The arrow indicates the splice junction between the last nucleotide of target βHCG6 exon I

15 and the first nucleotide encoding DT-A.

Figure 4A. Schematic diagram of the "safety" PTM and variations, demonstrating the PTM intramolecular base-paired stem, intended to mask the BP and PPT from splicing factors. Underlined sequences represent the βHCG6 intron 1 complementary target-binding domain, sequence in italics indicate target mismatches

20 that are homologous to the BP.

Figure 4B. Schematic of a safety PTM in open configuration upon binding to the target.

Figure 4C. *In vitro trans*-splicing reactions were carried out by incubating either safety PTM or safety PTM variants with the βHCG6 target. Splicing

25 reactions were amplified by RT-PCR using βHCG-F and DT-3R primers; products were analyzed in a 2.0% agarose gel.

Figure 5. Specificity of targeted *trans*-splicing is enhanced by the inclusion of a safety into the PTM. βHCG6 pre-mRNA (250 ng) and β-globin pre-mRNA (250 ng) were annealed together with either PTM+SF (safety) or pcPTM+Sp

30 (linear) RNA (500 ng). *In vitro trans*-splicing reactions and RT-PCR analysis were

8

performed as described under experimental procedures and the products were separated on a 2.0% agarose gel. Primers used for RT-PCR are as indicated.

Figure 6. In the presence of increasing PTM concentration, *cis*-splicing is inhibited and replaced by *trans*-splicing. *In vitro* splicing reactions were performed in the presence of a constant amount of βHCG6 target pre-mRNA (100 ng) with increasing concentrations of PTM (pcPTM+Sp) RNA (52-300 ng). RT-PCR for *cis*-spliced and un-spliced products utilized primers βHCG-F (exon 1 specific) and βHCG-R2 (exon 2 specific - Panel A); primers βHCG-F and DT-3R were used to RT-PCR *trans*-spliced products (Panel B). Reaction products were analyzed on 1.5% and 2.0% agarose gels, respectively. In panel A, lane 9 represents the 60 min time point in the presence of 300 ng of PTM, which is equivalent to lane 10 in panel B.

Figure 7A. PTMs are capable of *trans*-splicing in cultured human cancer cells. Total RNA was isolated from each of 4 expanded neomycin resistant H1299 lung carcinoma colonies transfected with pcSp+CRM (expressing non-toxic mutant DT-A) RT-PCR was performed using 1 µg of total RNA and 5' biotinylated βHCG-F and non-biotinylated DT-3R primers. Single stranded DNA was purified and sequenced.

Figure 7B. Nucleotide sequence (sense strand) of the *trans*-spliced product between endogenous βHCG6 target and CRM197 mutant toxin is shown. Two arrows indicate the position of the splice junction.

Figure 8A. Schematic diagram of a double splicing pre-therapeutic mRNA.

Figure 8B. Selective *trans*-splicing of a double splicing PTM. By varying the PTM concentration the PTM can be *trans*-spliced into either the 5' or the 3' splice site of the target.

Figure 9. Schematic diagram of the use of PTM molecules for exon tagging. Two examples of PTMs are shown. The PTM on the left is capable of non-specifically trans-splicing into a target pre-mRNA 3' splice site. The other PTM on the right is designed to non-specifically trans-splice into a target pre-mRNA 5' splice site. A PTM mediated trans-splicing reaction will result in the production of a chimeric RNA comprising a specific tag to either the 5' or 3' side of an authentic exon.

Figure 10A. Schematic diagram of constructs for use in the lacZ knock-out model. The target lacZ pre-mRNA contains the 5' fragment of lacZ followed by βHCG6 intron 1 and the 3' fragment of lacZ (target 1). The PTM molecule for use in the model system was created by digesting pPTM +SP with PstI and HindIII and replacing the DT-A toxin with βHCG6 exon 2 (pc3.1PTM2).

Figure 10B. Schematic diagram of restoration of β-Gal activity by Splicesome Mediated RNA Trans-splicing (SMaRT).

Figure 10C. Schematic diagram of constructs for use in the lacZ knock-in model (pc.3.1 lacZ T2). The lacZ target pre-mRNA is identical to that target pre-mRNA used for the knock-out experiments except that it contains two stop codons (TAA TAA) in frame four codons after the 3' splice site. The PTM molecule for use in the model system was created by digesting pPTM +SP with PstI and HindIII and replacing the DT-A toxin with functional 3' fragment of lacZ.

Figure 11A. Demonstration of cis-and trans-splicing when utilizing the lacZ knock-out model. The lacZ splice target 1 pre-mRNA and PTM2 were co-transfected into 293T cells. Total RNA was then isolated and analyzed by PCR for cis-spliced and trans-spliced products using the appropriate specific primers. The amplified PCR products were separated on a 2% agarose gel.

Figure 11B-C. Assays for β-galactosidase activity. 293 cells were transfected with lacZ target 2 DNA alone (panel B) or lacZ target 2 DNA and PTM1 (panel C).

Figure 12A. Nucleotide sequence of trans-spliced molecule demonstrating accurate trans-splicing.

Figure 12B. Nucleotide sequences of the cis-spliced product and the trans-spliced product. The nucleotide sequences were those sequences expected for each of the different splicing reactions.

Figure 13. Gene repair model for repair of the cystic fibrosis transmembrane regulator (CFTR) gene.

Figure 14. RT-PCR demonstration of trans-splicing between an exogenously supplied CFTR mini-gene target and PTM. Plasmids were co-transfected into 293 embryonic kidney cells. The primers pairs used for RT-PCR

10

reactions are listed above each lane. The lower band (471 bp) in each lane represents a *trans*-spliced product. The lower band in lane 1 (471bp) was purified from a 2% Seakem agarose gel and the DNA sequence of the band was determined.

Figure 15. DNA sequence of the *trans*-spliced product (lane 1, lower band shown in Figure 14). The DNA sequence indicates the presence of the F508 codon (TTT), exon 9 sequence is contiguous with exon 10 sequence, and the His tag sequence.

Figure 16. Schematic representation of repair of an exogenously supplied CFTR target molecule carrying an F508 deletion in exon 10.

Figure 17. Repair of endogenous CFTR transcripts by exon 10 invasion using a double splicing PTM. The use of a double splicing PTM permits repair of the Δ508 mutation with a very short PTM molecule.

5.   DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to compositions comprising pre-*trans*-splicing molecules (PTMs) and the use of such molecules for generating novel nucleic acid molecules. The PTMs of the invention comprise a target binding domain that is designed to specifically bind to pre-mRNA, a 3' splice region that includes a branch point, pyrimidine tract and a 3' splice acceptor site and/or a 5' splice donor site; and a spacer region that separates the RNA splice site from the target binding domain. In addition, the PTMs of the invention can be engineered to contain any nucleotide sequences encoding a translatable protein product.

The methods of the invention encompass contacting the PTMs of the invention with a natural pre-mRNA under conditions in which a portion of the PTM is *trans*-spliced to a portion of the natural pre-mRNA to form a novel chimeric RNA. The target pre-mRNA is chosen as a target due to its expression within a specific cell type thus providing a mechanism for targeting expression of a novel RNA to a selected cell type. The resulting chimeric RNA may provide a desired function, or may produce a gene product in the specific cell type. The specific cells may include, but are not limited to those infected with viral or other infectious agents, benign or malignant neoplasms, or components of the immune system which are involved in

11

autoimmune disease or tissue rejection. Specificity is achieved by modification of the
binding domain of the PTM to bind to the target endogenous pre-mRNA. The gene
products encoded by the chimeric RNA can be any gene, including genes having
clinical usefulness, for example, therapeutic or marker genes, and genes encoding

5   toxins.

5.1. STRUCTURE OF THE PRE-*TRANS*-SPLICING MOLECULES

The present invention provides compositions for use in generating
novel chimeric nucleic acid molecules through targeted *trans*-splicing. The PTMs of
the invention comprise (i) one or more target binding domains that targets binding of

10   the PTM to a pre-mRNA (ii) a 3' splice region that includes a branch point, pyrimidine
tract and a 3' splice acceptor site and/or 5' splice donor site; and (iii) a spacer region to
separate the RNA splice site from the target binding domain. Additionally, the PTMs
can be engineered to contain any nucleotide sequence encoding a translatable protein
product. In yet another embodiment of the invention, the PTMs can be engineered to

15   contain nucleotide sequences that inhibit the translation of the chimeric RNA
molecule. For example, the nucleotide sequences may contain translational stop
codons or nucleotide sequences that form secondary structures and thereby inhibit
translation. Alternatively, the chimeric RNA may function as an antisense molecule
thereby inhibiting translation of the RNA to which it binds.

20   The target binding domain of the PTM may contain one or two binding
domains of at least 15 to 30 (up to several hundred) nucleotides which are
complementary to and in anti-sense orientation to the targeted region of the selected
pre-mRNA. This confers specificity of binding and anchors the pre-mRNA closely in
space so that the spliceosome processing machinery of the nucleus can *trans*-splice a

25   portion of the PTM to a portion of the pre-mRNA. A second target binding region
may be placed at the 3' end of the molecule and can be incorporated into the PTM of
the invention. Absolute complementarity, although preferred, is not required. A
sequence "complementary" to a portion of an RNA, as referred to herein, means a
sequence having sufficient complementarity to be able to hybridize with the RNA,

30   forming a stable duplex. The ability to hybridize will depend on both the degree of

12

complementarity and the length of the nucleic acid (See, for example, Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York). Generally, the longer the hybridizing nucleic acid, the more base mismatches with an RNA it may contain and still form a stable duplex. One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures to determine the melting point of the hybridized complex.

Where the PTMs are designed for use in intron-exon tagging or for peptide affinity tagging, a library of PTMs is genetically engineered to contain random nucleotide sequences in the target binding domain. The goal of generating such a library of PTM molecules is that the library will contain a population of PTM molecules capable of binding to each RNA molecule expressed in the cell. A recombinant expression vector can be genetically engineered to contain a coding region for a PTM including a restriction endonuclease site that can be used for insertion of random DNA fragments into the PTM to form random target binding domains. The random nucleotide sequences to be included in the PTM as target binding domains can be generated using a variety of different methods well known to those of skill in the art, including but not limited to, partial digestion of DNA with restriction enzymes or mechanical shearing of DNA to generate random fragments of DNA. Random binding domain regions may also be generated by degenerate oligonucleotide synthesis. The degenerate oligonucleotides can be engineered to have restriction endonuclease recognition sites on each end to facilitate cloning into a PTM molecule for production of a library of PTM molecules having degenerate binding domains.

Binding may also be achieved through other mechanisms, for example, through triple helix formation or protein/nucleic acid interactions such as those in which the PTM is engineered to recognize a specific RNA binding protein, i.e., a protein bound to a specific target pre-mRNA. Alternatively, the PTMs of the invention may be designed to recognize secondary structures, such as for example, hairpin structures resulting from intramolecular base pairing between nucleotides within an RNA molecule.

13

The PTM molecule also contains a 3' splice region that includes a branch point, pyrimidine tract and a 3' splice acceptor AG site and/or a 5' splice donor site. Consensus sequences for the 5' splice donor site and the 3' splice region used in RNA splicing are well known in the art (See, Moore, *et al.*, 1993, The RNA World, Cold Spring Harbor Laboratory Press, p. 303-358). In addition, modified consensus sequences that maintain the ability to function as 5' donor splice sites and 3' splice regions may be used in the practice of the invention. Briefly, the 5' splice site consensus sequence is AG/GURAGU (where A=adenosine, U=uracil, G=guanine, C=cytosine, R=purine and /=the splice site). The 3' splice site consists of three separate sequence elements: the branch point or branch site, a polypyrimidine tract and the 3' consensus sequence (YAG). The branch point consensus sequence in mammals is YNYURAC (Y=pyrimidine). The underlined A is the site of branch formation. A polypyrimidine tract is located between the branch point and the splice site acceptor and is important for different branch point utilization and 3' splice site recognition.

A spacer region to separate the RNA splice site from the target binding domain is also included in the PTM. The spacer region can have features such as stop codons which would block any translation of an unspliced PTM and/or sequences that enhance *trans*-splicing to the target pre-mRNA.

In a preferred embodiment of the invention, a "safety" is also incorporated into the spacer, binding domain, or elsewhere in the PTM to prevent non-specific *trans*-splicing. This is a region of the PTM that covers elements of the 3' and/or 5' splice site of the PTM by relatively weak complementarity, preventing non-specific *trans*-splicing. The PTM is designed in such a way that upon hybridization of the binding /targeting portion(s) of the PTM, the 3' and/or 5'splice site is uncovered and becomes fully active.

The "safety" consists of one or more complementary stretches of *cis*-sequence (or could be a second, separate, strand of nucleic acid) which weakly binds to one or both sides of the PTM branch point, pyrimidine tract, and/or 3' splice site (splicing elements), or could bind to parts of the splicing elements themselves. This "safety" binding prevents the splicing elements from being active (i.e. block U2

14

snRNP or other splicing factors from attaching to the PTM splice site recognition
elements). The binding of the "safety" may be disrupted by the binding of the target
binding region of the PTM to the target pre-mRNA, thus exposing and activating the
PTM splicing elements (making them available to *trans*-splice into the target pre-
5   mRNA).

A nucleotide sequence encoding a translatable protein capable of
producing an effect, such as cell death, or alternatively, one that restores a missing
function or acts as a marker, is included in the PTM of the invention. For example,
the nucleotide sequence can include those sequences encoding gene products missing
10   or altered in known genetic diseases. Alternatively, the nucleotide sequences can
encode marker proteins or peptides which may be used to identify or image cells. In
yet another embodiment of the invention nucleotide sequences encoding affinity tags
such as, HIS tags (6 consecutive histidine residues) (Janknecht, et al., 1991, Proc.
Natl. Acad. Sci. USA 88:8972-8976), the C-terminus of glutathione-S-transferase
15   (GST) (Smith and Johnson, 1986, Proc. Natl. Acad. Sci. USA 83:8703--8707)
(Pharmacia) or FLAG (Asp-Tyr-Lys-Asp-Asp-Asp-Lys) (Eastman Kodak/IBI,
Rochester, NY) can be included in PTM molecules for use in affinity purification.
The use of PTMs containing such nucleotide sequences results in the production of a
chimeric RNA encoding a fusion protein containing peptide sequences normally
20   expressed in a cell linked to the peptide affinity tag. The affinity tag provides a
method for the rapid purification and identification of peptide sequences expressed in
the cell. In a preferred embodiment the nucleotide sequences may encode toxins or
other proteins which provide some function which enhances the susceptibility of the
cells to subsequent treatments, such as radiation or chemotherapy.

25   In a highly preferred embodiment of the invention a PTM molecule is
designed to contain nucleotide sequences encoding the Diphtheria toxin subunit A
(Greenfield, L., et al., 1983, Proc. Nat'l. Acad. Sci. USA 80: 6853-6857). Diphtheria
toxin subunit A contains enzymatic toxin activity and will function if expressed or
delivered into human cells resulting in cell death. Furthermore, various other known
30   peptide toxins may be used in the present invention, including but not limited to, ricin,
Pseudomonus toxin, *Shiga* toxin and exotoxin A.

15

Additional features can be added to the PTM molecule either after, or before, the nucleotide sequence encoding a translatable protein, such as polyadenylation signals or 5' splice sequences to enhance splicing, additional binding regions, "safety"-self complementary regions, additional splice sites, or protective

5   groups to modulate the stability of the molecule and prevent degradation.

Additional features that may be incorporated into the PTMs of the invention include stop codons or other elements in the region between the binding domain and the splice site to prevent unspliced pre-mRNA expression. In another embodiment of the invention, PTMs can be generated with a second anti-sense

10   binding domain downstream from the nucleotide sequences encoding a translatable protein to promote binding to the 3' target intron or exon and to block the fixed authentic *cis*-5' splice site (U5 and/or U1 binding sites). PTMs may also be made that require a double *trans*-splicing reaction for expression of the *trans*-spliced product. Such PTMs could be used to replace an internal exon which could be useful for RNA

15   repair. Further elements such as a 3' hairpin structure, circularized RNA, nucleotide base modification, or a synthetic analog can be incorporated into PTMs to promote or facilitate nuclear localization and spliceosomal incorporation, and intra-cellular stability.

The PTMs of the invention can be used in methods designed to produce a novel chimeric RNA in a target cell. The methods of the present invention

20   comprise delivering to the target cell a PTM which may be in any form used by one skilled in the art, for example, an RNA molecule, or a DNA vector which is transcribed into a RNA molecule, wherein said PTM binds to a pre-mRNA and mediates a *trans*-splicing reaction resulting in formation of a chimeric RNA

25   comprising a portion of the PTM molecule spliced to a portion of the pre-mRNA.

5.2. SYNTHESIS OF THE *TRANS*-SPLICING MOLECULES

The nucleic acid molecules of the invention can be RNA or DNA or derivatives or modified versions thereof, single-stranded or double-stranded. By nucleic acid is meant a PTM molecule or a nucleic acid molecule encoding a PTM

30   molecule, whether composed of deoxyribonucleotides or ribonucleosides, and whether

16

composed of phosphodiester linkages or modified linkages. The term nucleic acid also specifically includes nucleic acids composed of bases other than the five biologically occurring bases (adenine, guanine, thymine, cytosine and uracil).

The RNA and DNA molecules of the invention can be prepared by any
5   method known in the art for the synthesis of DNA and RNA molecules. For example, the nucleic acids may be chemically synthesized using commercially available reagents and synthesizers by methods that are well known in the art (see, e.e., Gait, 1985, Oligonucleotide Synthesis: A Practical Approach, IRL Press, Oxford, England). Alternatively, RNA molecules can be generated by *in vitro* and *in vivo* transcription of
10   DNA sequences encoding the RNA molecule. Such DNA sequences can be incorporated into a wide variety of vectors which incorporate suitable RNA polymerase promoters such as the T7 or SP6 polymerase promoters. RNAs may be produced in high yield via *in vitro* transcription using plasmids such as SPS65 (Promega Corporation, Madison, WI). In addition, RNA amplification methods such
15   as Q-β amplification can be utilized to produce RNAs.

The nucleic acid molecules can be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule, hybridization, transport into the cell, etc. For example, modification of a PTM to reduce the overall charge can enhance the cellular uptake of the molecule. In addition
20   modifications can be made to reduce susceptibility to nuclease degradation. The nucleic acid molecules may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, 1989, Proc. Natl. Acad. Sci. U.S.A. 86:6553-6556; Lemaitre *et al.*, 1987, Proc. Natl. Acad. Sci. 84:648-652; PCT Publication No.
25   WO88/09810, published December 15, 1988) or the blood-brain barrier (see, *e.g.*, PCT Publication No. WO89/10134, published April 25, 1988), hybridization-triggered cleavage agents. (See, *e.g.*, Krol *et al.*, 1988, BioTechniques 6:958-976) or intercalating agents. (See, *e.g.*, Zon, 1988, Pharm. Res. 5:539-549). To this end, the nucleic acid molecules may be conjugated to another molecule, *e.g.*, a peptide,
30   hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc. Various other well-known modifications to the DNA molecules

17

can be introduced as a means of increasing intracellular stability and half-life.

Possible modifications include, but are not limited to, the addition of flanking

sequences of ribo- or deoxy- nucleotides to the 5' and/or 3' ends of the molecule. In

some circumstances where increased stability is desired, nucleic acids having

5   modified internucleoside linkages such as 2'-0-methylation may be preferred. Nucleic

acids containing modified internucleoside linkages may be synthesized using reagents

and methods that are well known in the art (see, Uhlmann *et al.*, 1990, Chem. Rev.

90:543-584; Schneider *et al.*, 1990, Tetrahedron Lett. 31:335 and references sited

therein).

10   The nucleic acids may be purified by any suitable means, as are well

known in the art. For example, the nucleic acids can be purified by reverse phase

chromatography or gel electrophoresis. Of course, the skilled artisan will recognize

that the method of purification will depend in part on the size of the nucleic acid to be

purified.

15   In instances where a nucleic acid molecule encoding a PTM is utilized,

cloning techniques known in the art may be used for cloning of the nucleic acid

molecule into an expression vector. Methods commonly known in the art of

recombinant DNA technology which can be used are described in Ausubel *et al.*

(eds.), 1993, Current Protocols in Molecular Biology, John Wiley & Sons, NY; and

20   Kriegler, 1990, Gene Transfer and Expression, A Laboratory Manual, Stockton Press,

NY.

              The DNA encoding the PTM of interest may be recombinantly

engineered into a variety of host vector systems that also provide for replication of the

DNA in large scale and contain the necessary elements for directing the transcription

25   of the PTM. The use of such a construct to transfect target cells in the patient will

result in the transcription of sufficient amounts of PTMs that will form

complementary base pairs with the endogenously expressed pre-mRNA targets and

thereby facilitate a *trans*-splicing reaction between the complexed nucleic acid

molecules. For example, a vector can be introduced *in vivo* such that it is taken up by

30   a cell and directs the transcription of the PTM molecule. Such a vector can remain

episomal or become chromosomally integrated, as long as it can be transcribed to

18

produce the desired RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art.

Vectors encoding the PTM of interest can be plasmid, viral, or others known in the art, used for replication and expression in mammalian cells. Expression

5   of the sequence encoding the PTM can be regulated by any promoter known in the art to act in mammalian, preferably human cells. Such promoters can be inducible or constitutive. Such promoters include but are not limited to: the SV40 early promoter region (Benoist, C. and Chambon, P. 1981, Nature 290:304-310), the promoter contained in the 3' long terminal repeat of Rous sarcoma virus (Yamamoto *et al.*,

10   1980, Cell 22:787-797), the herpes thymidine kinase promoter (Wagner *et al.*, 1981, Proc. Natl. Acad. Sci. U.S.A. 78:14411445), the regulatory sequences of the metallothionein gene (Brinster *et al.*, 1982, Nature 296:39-42), the viral CMV promoter, the human chorionic gonadotropin-β promoter (Hollenberg et al., 1994, Mol. Cell. Endocrinology 106:111-119), etc. Any type of plasmid, cosmid, YAC or

15   viral vector can be used to prepare the recombinant DNA construct which can be introduced directly into the tissue site. Alternatively, viral vectors can be used which selectively infect the desired target cell.

For use of PTMs encoding peptide affinity purification tags, it is desirable to insert nucleotide sequences containing random target binding sites into

20   the PTMs and clone them into a selectable mammalian expression vector system. A number of selection systems can be used, including but not limited to selection for expression of the herpes simplex virus thymidine kinase, hypoxanthine-guanine phosphoribosyltransterase and adenine phosphoribosyl tranferase protein in tk-, hgprt- or aprt- deficient cells, respectively. Also, anti-metabolic resistance can be used as the

25   basis of selection for dihydrofolate tranferase (*dhfr*), which confers resistance to methotrexate; xanthine-guanine phosphoribosyl transferase *(gpt)*, which confers resistance to mycophenolic acid; neomycin (*neo*), which confers resistance to aminoglycoside G-418; and hygromycin B phosphotransferase (*hygro*) which confers resistance to hygromycin. In a preferred embodiment of the invention, the cell culture

30   is transformed at a low ratio of vector to cell such that there will be only a single vector, or a limited number of vectors, present in any one cell. Vectors for use in the

19

practice of the invention include any eukaryotic expression vectors, including but not limited to viral expression vectors such as those derived from the class of retroviruses or adeno-associated viruses.

5.3. USES AND ADMINISTRATION OF *TRANS*-SPLICING MOLECULES

5   5.3.1   USE OF PTM MOLECULES FOR GENE REGULATION, GENE REPAIR AND TARGETED CELL DEATH

The compositions and methods of the present invention will have a variety of different applications including gene regulation, gene repair and targeted cell death. For example, *trans*-splicing can be used to introduce a protein with toxic

10   properties into a cell. In addition, PTMs can be engineered to bind to viral mRNA and destroy the function of the viral mRNA, or alternatively, to destroy any cell expressing the viral mRNA. In yet another embodiment of the invention, PTMs can be engineered to place a stop codon in a deleterious mRNA transcript thereby decreasing the expression of that transcript.

15   Targeted *trans*-splicing can be used to repair or correct transcripts that are either truncated or contain point mutations. The PTMs of the invention are designed to cleave a targeted transcript upstream or downstream of a specific mutation or upstream of a premature 3' and correct the mutant transcript via a *trans*-splicing reaction which replaces the portion of the transcript containing the mutation with a

20   functional sequence.

Cystic fibrosis (CF) is the most common fatal genetic disease in humans. Based on both genetic and molecular analyses, the gene associated with cystic fibrosis has been isolated and its protein product deduced (Kerem, B.S. et al., 1989, Science 245:1073-1080; Riordan et al., 1989, Science 245:1066-

25   1073;Rommans, et al., 1989, Science 245:1059-1065). The protein product of the CF associated gene is called the cystic fibrosis transmembrane conductance regulator (CFTR). In a specific embodiment of the invention, a *trans*-splicing reaction will be used to correct a genetic defect in the DNA sequence encoding the cystic fibrosis transmembrane regulator (CFTR) whereby the DNA sequence encoding the cystic

20

fibrosis trans-membrane regulator protein is expressed and a functional chloride ion channel is produced in the airway epithelial cells of a patient.

Population studies have indicated that the most common cystic fibrosis mutation is a deletion of the three nucleotides in exon 10 that encode phenylalanine at position 508 of the CFTR amino acid sequence. As indicated in Figure 15, a *trans*-splicing reaction was capable of correcting the deletion at position 508 in the CFTR amino acid sequence. The PTM used for correction of the genetic defect contained a CFTR BD intron 9 sequence, a spacer sequence, a branch point, a polypyrimidine tract, a 3' splice site and a wild type CFTR BD exon 10 sequence (Figure 13). The successful correction of the mutated DNA encoding CFTR utilizing a *trans*-splicing reaction supports the general application of PTMs for correction of genetic defects.

Various delivery systems are known and can be used to transfer the compositions of the invention into cells, *e.g.* encapsulation in liposomes, microparticles, microcapsules, recombinant cells capable of expressing the composition, receptor-mediated endocytosis (see, *e.g.*, Wu and Wu, 1987, J. Biol. Chem. 262:4429-4432), construction of a nucleic acid as part of a retroviral or other vector, injection of DNA, electroporation, calcium phosphate mediated transfection, etc.

The compositions and methods can be used to treat cancer and other serious viral infections, autoimmune disorders, and other pathological conditions in which the alteration or elimination of a specific cell type would be beneficial. Additionally, the compositions and methods may also be used to provide a gene encoding a functional biologically active molecule to cells of an individual with an inherited genetic disorder where expression of the missing or mutant gene product produces a normal phenotype.

In a preferred embodiment, nucleic acids comprising a sequence encoding a PTM are administered to promote PTM function, by way of gene delivery and expression into a host cell. In this embodiment of the invention, the nucleic acid mediates an effect by promoting PTM production. Any of the methods for gene delivery into a host cell available in the art can be used according to the present invention. For general reviews of the methods of gene delivery see Strauss, M. and

21

Barranger, J.A., 1997, Concepts in Gene Therapy, by Walter de Gruyter & Co., Berlin;

Goldspiel *et al.*, 1993, Clinical Pharmacy 12:488-505; Wu and Wu, 1991, Biotherapy

3:87-95; Tolstoshev, 1993, Ann. Rev. Pharmacol. Toxicol. 33:573-596; Mulligan,

1993, Science 260:926-932; and Morgan and Anderson, 1993, Ann. Rev. Biochem.

5   62:191-217; 1993, TIBTECH 11(5):155-215. Exemplary methods are described

below.

Delivery of the nucleic acid into a host cell may be either direct, in

which case the host is directly exposed to the nucleic acid or nucleic acid-carrying

vector, or indirect, in which case, host cells are first transformed with the nucleic acid

10   *in vitro*, then transplanted into the host. These two approaches are known,

respectively, as *in vivo* or *ex vivo* gene delivery.

In a specific embodiment, the nucleic acid is directly administered *in*

vivo, where it is expressed to produce the PTM. This can be accomplished by any of

numerous methods known in the art, *e.g.*, by constructing it as part of an appropriate

15   nucleic acid expression vector and administering it so that it becomes intracellular,

e.g. by infection using a defective or attenuated retroviral or other viral vector (see

U.S. Patent No. 4,980,286), or by direct injection of naked DNA, or by use of

microparticle bombardment (*e.g.*, a gene gun; Biolistic, Dupont), or coating with

lipids or cell-surface receptors or transfecting agents, encapsulation in liposomes,

20   microparticles, or microcapsules, or by administering it in linkage to a peptide which

is known to enter the nucleus, by administering it in linkage to a ligand subject to

receptor-mediated endocytosis (see *e.g.*, Wu and Wu, 1987, J. Biol. Chem. 262:4429-

4432).

In a specific embodiment, a viral vector that contains the PTM can be

25   used. For example, a retroviral vector can be utilized that has been modified to delete

retroviral sequences that are not necessary for packaging of the viral genome and

integration into host cell DNA (see Miller *et al.*, 1993, Meth. Enzymol. 217:581-599).

Alternatively, adenoviral or adeno-associated viral vectors can be used for gene

delivery to cells or tissues. (See, Kozarsky and Wilson, 1993, Current Opinion in

30   Genetics and Development 3:499-503 for a review of adenovirus-based gene

delivery).

22

Another approach to gene delivery into a cell involves transferring a gene to cells in tissue culture by such methods as electroporation, lipofection, calcium phosphate mediated transfection, or viral infection. Usually, the method of transfer includes the transfer of a selectable marker to the cells. The cells are then placed under selection to isolate those cells that have taken up and are expressing the transferred gene. The resulting recombinant cells can be delivered to a host by various methods known in the art. In a preferred embodiment, the cell used for gene delivery is autologous to the host cell.

The present invention also provides for pharmaceutical compositions comprising an effective amount of a PTM or a nucleic acid encoding a PTM, and a pharmaceutically acceptable carrier. In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, and more particularly in humans. The term "carrier" refers to a diluent, adjuvant, excipient, or vehicle with which the therapeutic is administered. Examples of suitable pharmaceutical carriers are described in "Remington's Pharmaceutical sciences" by E.W. Martin.

In specific embodiments, pharmaceutical compositions are administered: (1) in diseases or disorders involving an absence or decreased (relative to normal or desired) level of an endogenous protein or function, for example, in hosts where the protein is lacking, genetically defective, biologically inactive or underactive, or under expressed; or (2) in diseases or disorders wherein, *in vitro* or *in vivo*, assays indicate the utility of PTMs that inhibit the function of a particular protein . The activity of the protein encoded for by the chimeric mRNA resulting from the PTM mediated *trans*-splicing reaction can be readily detected, *e.g.*, by obtaining a host tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for mRNA or protein levels, structure and/or activity of the expressed chimeric mRNA. Many methods standard in the art can be thus employed, including but not limited to immunoassays to detect and/or visualize the protein encoded for by the chimeric mRNA (*e.g.*, Western blot, immunoprecipitation followed by sodium dodecyl sulfate polyacrylamide gel electrophoresis, immunocytochemistry, etc.) and/or hybridization

assays to detect formation of chimeric mRNA expression by detecting and/or visualizing the presence of chimeric mRNA (*e.g.*, Northern assays, dot blots, in situ hybridization, and Reverse-Transcription PCR, etc.), etc.

The present invention also provides for pharmaceutical compositions comprising an effective amount of a PTM or a nucleic acid encoding a PTM, and a pharmaceutically acceptable carrier. In a specific embodiment, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, and more particularly in humans. The term "carrier" refers to a diluent, adjuvant, excipient, or vehicle with which the therapeutic is administered. Examples of suitable pharmaceutical carriers are described in "Remington's Pharmaceutical sciences" by E.W. Martin. In a specific embodiment, it may be desirable to administer the pharmaceutical compositions of the invention locally to the area in need of treatment. This may be achieved by, for example, and not by way of limitation, local infusion during surgery, topical application, *e.g.*, in conjunction with a wound dressing after surgery, by injection, by means of a catheter, by means of a suppository, or by means of an implant, said implant being of a porous, non-porous, or gelatinous material, including membranes, such as sialastic membranes, or fibers. Other control release drug delivery systems, such as nanoparticles, matrices such as controlled-release polymers, hydrogels.

The PTM will be administered in amounts which are effective to produce the desired effect in the targeted cell. Effective dosages of the PTMs can be determined through procedures well known to those in the art which address such parameters as biological half-life, bioavailability and toxicity. The amount of the composition of the invention which will be effective will depend on the nature of the disease or disorder being treated, and can be determined by standard clinical techniques. In addition, *in vitro* assays may optionally be employed to help identify optimal dosage ranges.

The present invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention optionally associated with such

24

container(s) can be a notice in the form prescribed by a governmental agency
regulating the manufacture, use or sale of pharmaceuticals or biological products,
which notice reflects approval by the agency of manufacture, use or sale for human
administration.

5   5.3.2   USE OF PTM MOLECULES FOR EXON TAGGING
                In view of current efforts to sequence and characterize the genomes of
humans and other organisms, there is a need for methods that facilitate such
characterization.   A majority of the information currently obtained by genomic
mapping and sequencing is derived from complementary DNA (cDNA) libraries,
10   which are made by reverse transcription of mRNA into cDNA. Unfortunately, this
process causes the loss of information concerning intron sequences and the location of
exon/intron boundaries.
                The present invention encompasses a method for mapping exon-intron
boundaries in pre-mRNA molecules comprising (i) contacting a pre-transplicing
15   molecule containing a random target binding domain with a pre-mRNA molecule
under conditions in which a portion of the pre-trans-splicing molecule is trans-spliced
to a portion of the target pre-mRNA to form a chimeric mRNA; (ii) amplifying the
chimeric mRNA molecule; (iii) selectively purifying the amplified molecule; and (iv)
determining the nucleotide sequence of the amplified molecule thereby identifying the
20   intron-exon boundaries.
                In an embodiment of the present invention, PTMs can be used in trans-
splicing reactions to locate exon-intron boundaries in pre-mRNAs molecules. PTMs
for use in mapping of intron-exon boundaries have structures similar to those
described above in Section 5.1. Specifically, the PTMs contain (i) a target binding
25   domain that is designed to bind to many pre-mRNAs: (ii) a 3' splice region that
includes a branch point, pyrimidine tract and a 3' splice acceptor site, or a 5' splice
donor site; (iii) a spacer region that separates the mRNA splice site from the target
binding domain; and (iv) a tag region that will be trans-spliced onto a pre-mRNA. For
purposes of intron-exon mapping, the PTMs are genetically engineered to contain
30   target binding domains comprising random nucleotide sequences. The random

25

nucleotide sequences contain at least 15-30 and up to several hundred nucleotide

sequences capable of binding and anchoring a pre-mRNA so that the spliceosome

processing machinery of the nucleus can trans-splice a portion (tag or marker region)

of the PTM to a portion of the pre-mRNA. PTMs containing short target binding

5   domains, or containing inosines bind under less stringent conditions to the pre-mRNA

molecules. In addition, strong branch point sequences and pyrimidine tracts serve to

increase the non-specificity of PTM trans-splicing.

The random nucleotide sequences used as target binding domains in

the PTM molecules can be generated using a variety of different methods, including,

10   but not limited to, partial digestion of DNA with restriction endonucleases or

mechanical shearing of the DNA. The use of such random nucleotide sequences is

designed to generate a vast array of PTM molecules with different binding activities

for each target pre-mRNA expressed in a cell. Randomized libraries of

oligonucleotides can be synthesized with appropriate restriction endonucleases

15   recognition sites on each end for cloning into PTM molecules genetically engineered

into plasmid vectors. When the randomized oligonucleotides are litigated and

expressed, a randomized binding library of PTMs is generated.

In a specific embodiment of the invention, an expression library

encoding PTM molecules containing target binding domains comprising random

20   nucleotide sequences can be generated using a variety of methods which are well

known to those of skill in the art. Ideally, the library is complex enough to contain

PTM molecules capable of interacting with each target pre-mRNA expressed in a cell.

By way of example, Figure 9 is a schematic representation of two

forms of PTMs which can be utilized to map intron-exon boundaries. The PTM on

25   the left is capable of non-specifically trans-splicing into a pre-mRNA 3' splice site,

while the PTM on the right is capable of trans-splicing into a pre-mRNA 5' splice site.

Trans-splicing between the PTM and the target pre-mRNA results in the production of

a chimeric mRNA molecule having a specific nucleotide sequence "tag" on either the

3' or 5' end of an authentic exon.

30   Following selective purification, a DNA sequencing reaction is then

performed using a primer which begins in the tag nucleotide sequence of the PTM and

26

proceeds into the sequence of the tagged exon. The sequence immediately following

the last nucleotide of the tag nucleotide sequence represents an exon boundary. For

identification of intron-exon tags, the trans-splicing reactions of the invention can be

performed either *in vitro* or *in vivo* using methods well known to those of skill in the

5   art.

5.3.3   USE OF PTM MOLECULES FOR IDENTIFICATION OF PROTEINS EXPRESSED IN A CELL

In yet another embodiment of the invention, PTM mediated trans-

splicing reactions can be used to identify previously undetected and unknown proteins

10   expressed in a cell. This method is especially useful for identification of proteins that

cannot be detected by a two-dimensional electrophoresis, or by other methods, due to

inter alia the small size of the protein, low concentration of the protein, or failure to

detect the protein due to similar migration patterns with other proteins in two-

dimensional electrophoresis.

15   The present invention relates to a method for identifying proteins

expressed in a cell comprising (i) contacting a pre-transplicing molecule containing a

random target binding domain and a nucleotide sequence encoding a peptide tag with

a pre-mRNA molecule under conditions in which a portion of the pre-trans-splicing

molecule is trans-spliced to a portion of the target pre-mRNA to form a chimeric

20   mRNA encoding a fusion polypeptide or separating it by gel electrophoresis (ii)

affinity purifying the fusion polypeptide; and (iii) determining the amino acid

sequence of the fusion protein.

To identify proteins expressed in a cell, the PTMs of the invention are

genetically engineered to contain: (i) a target binding domain comprising randomized

25   nucleotide sequences; (ii) a 3' splice region that includes a branch point, pyrimidine

tract and a 3' splice acceptor site and/or a 5' splice donor site; (iii) a spacer region that

separates the PTM splice site from the target binding domain; and (iv) nucleotide

sequences encoding a marker or peptide affinity purification tag. Such peptide tags

include, but are not limited to, HIS tags (6 histidine consecutive residues) (Janknecht,

30   et al., 1991 Proc. Natl. Acad. Sci. USA 88:8972-8976), glutathione-S-transferase

27

(GST) (Smith, D.B. and Johnson K.S., 1988, Gene 67:31) (Pharmacia) or FLAG
(Kodak/IBI) tags (Nisson, J. et al. J. Mol. Recognit., 1996, 5:585-594)

Trans-splicing reactions using such PTMs results in the generation of
chimeric mRNA molecules encoding fusion proteins comprising protein sequences
normally expressed in a cell linked to a marker or peptide affinity purification tag. The
desired goal of such a method is that every protein synthesized in a cell receives a
marker or peptide affinity tag thereby providing a method for identifying each protein
expressed in a cell.

In a specific embodiment of the invention, PTM expression libraries
encoding PTMs having different target binding domains comprising random
nucleotide sequences are generated. The desired goal is to create a PTM expression
library that is complex enough to produce a PTM capable of binding to each pre-
mRNA expressed in a cell. In a preferred embodiment, the library is cloned into a
mammalian expression vector that results in one, or at most, a few vectors being
present in any one cell.

To identify the expression of chimeric proteins, host cells are
transformed with the PTM library and plated so that individual colonies containing
one PTM vector can be grown and purified. Single colonies are selected, isolated, and
propagated in the appropriate media and the labeled chimeric protein exon(s)
fragments are separated away from other cellular proteins using, for example, an
affinity purification tag. For example, affinity chromatography can involve the use of
antibodies that specifically bind to a peptide tag such as the FLAG tag. Alternatively,
when utilizing HIS tags, the fusion proteins are purified using a Ni^{2+} nitriloacetic acid
agarose columns, which allows selective elution of bound peptide eluted with
imidazole containing buffers. When using GST tags, the fusion proteins are purified
using glutathione-S-transferase agarose beads. The fusion proteins can then be eluted
in the presence of free glutathione.

Following purification of the chimeric protein, an analysis is carried
out to determine the amino acid sequence of the fusion protein. The amino acid
sequence of the fusion protein is determined using techniques well known to those of
skill in the art, such as Edman Degradation followed by amino acid analysis using

28

HPLC, mass spectrometry or an amino acid analyzation. Once identified, the peptide

sequence is compared to those sequences available in protein databases, such as

GenBank. If the partial peptide sequence is already known, no further analysis is

done. If the partial protein sequence is unknown, then a more complete sequence of

5   that protein can be carried out to determine the full protein sequence. Since the

fusion protein will contain only a portion of the full length protein, a nucleic acid

encoding the full length protein can be isolated using conventional methods. For

example, based on the partial protein sequence oligonucleotide primers can be

generated for use as probes or PCR primers to screen a cDNA library.

10

  6. EXAMPLE: PRODUCTION OF *TRANS*-SPLICING MOLECULES

    The following section describes the production of PTMs and the

demonstration that such molecules are capable of mediating *trans*-splicing reactions

resulting in the production of chimeric mRNA molecules.

15

    6.1 MATERIALS AND METHODS

    6.1.1. CONSTRUCTION OF PRE-mRNA MOLECULES

    Plasmids containing the wild type diphtheria toxin subunit A (DT-A,

wild-type accession #K01722) and a DT-A mutant (CRM 197, no enzymatic activity)

were obtained from Dr. Virginia Johnson, Food and Drug Administration, Bethesda,

20   Maryland (Uchida *et al.*, 1973 J. Biol. Chem 248:3838). For *in vitro* experiments,

DT-A was amplified using primers: DT-1F (5'-

GGCGCTGCAGGGCGCTGATGATGTTGTTG); and DT-2R (5'-GGCGAAG

CTTGGATCCGACACGATTTCCTGCACAGG), cut with PstI and HindIII, and

cloned into PstI and HindIII digested pBS(-) vector (Stratagene, La Jolla, CA). The

25   resulting clone, pDTA was used to construct the individual PTMs. (1) pPTM+:

Targeted construct. Created by inserting IN3-1 (5'AATTCTCTAGATGCTT

CACCCGGGCCTGACTCGAGTACTAACTGGTACCTCTTCTTTTTTTTCCTGCA

) and IN2-4 (5'-GGAAAAAAAAGAAGAGGTACCAGTTAGTACTCGAGTCAGG

CCCGGGTGAAGCATCTAGAG) primers into EcoRI and PstI digested pDTA. (2)

30   pPTM+Sp: As pPTM+ but with a 30 bp spacer sequence between the BD and BP.

29

Created by digesting pPTM+ with XhoI and ligating in the oligonucleotides, spacer S
(5'-TCGAGCAACGTTATAATAATGTTC) and spacer AS (5'-
TCGAGAACATTATT ATAACGTTGC). For *in vivo* studies, an EcoRI and HindIII
fragment of pcPTM+Sp was cloned into mammalian expression vector pcDNA3.1

5   (Invitrogen), under the control of a CMV promoter. Also, the methionine at codon 14
was changed into isoleucine to prevent initiation of translation. The resulting plasmid
was designated as pcPTM+Sp. (3) pPTM+CRM: As pPTM+Sp but the wild type DT-
A was substituted with CRM mutant DT-A (T. Uchida, et al., 1973, J. Biol. Chem.
248:3838). This was created by PCR amplification of a DT-A mutant (mutation at

10   G52E) using primers DT-1F and DT-2R. For *in vivo* studies, an EcoRI HindIII
fragment of PTM+CRM was cloned into pc3.1DNA that resulted in pcPTM+ARM.
(4) PTM-: Non-targeted construct. Created by digestion of PTM+ with EcoRI and Pst
I, gel purified to remove the binding domain followed by ligation of the
oligonucleotides, IN-5 (5'-ATCTCTAGATCAGGCCCGGGTGAAGCC CGAG) and

15   IN-6 (5'-TGCTTCACCC GGGCCTGATCTAGAG). (5) PTM-Sp, is an identical
version of the PTM-, except it has a 30 bp spacer sequence at the PstI site. Similarly,
the splice mutants [Py(-)AG(-) and BP(-)Py(-)AG(-)] and safety variants [PTM+SF-
Py1, PTM+SF-Py2, PTM+SFBP3 and PTM+SFBP3-Py1] were constructed either by
insertion or deletion of specific sequences (see Table 1).

PTM construct	BD/NBD	BP	PPT	3'ss
	Table 1. Binding/non-binding domain, BP, PPT and 3' as sequences of different PTMs.			
PTM+Sp (targeted)	:TGCTTCACCCGGGCCTGA	TACTAAC	CTCTTCTTTTTTTTCC	CAG
PTM-Sp (non-targeted)	:CAACGTTATAATAATGTT	TACTAAC	CTCTTCTTTTTTTTCC	CAG
PTM+Py (-)AG(-)BP(-)	:TGCTTCACCCGGGCCTGA	GGCTGAT	CTGTGATTAATAGCGG	ACG
PTM+Py(-)AG(-)	:TGCTTCACCCGGGCCTGA	TACTAAC	CCTGGACGCGGAAGTT	ACG
PTM+SF	:CTGGGACAAGGACACTGCTTCACCCGGTTAGTAGACCACAGCCCTGAAGCC	TACTAAC	CTTCTGTTTTTTTCTC	CAG
PTM+SF-Py	:As in PTM+SF	TACTAAC	CTTCTGTATTATTCTC	CAG
PTM+SF-Py	:As in PTM+SF	TACTAAC	GTTCTGTCCTTGTCTC	CAG
PTM+SF-BP3	:As in PTM+SF	TGCTGAC	CTTCTGTTTTTTTCTC	CAG
PTM+SFBP3-Py	:As in PTM+SF	TGCTGAC	CTTCTGTATTATTCTC	CAG

Nucleotides in bold indicate the mutations compared to normal BP, PPT and 3' splice site.

Branch site A is underlined. The nucleotides in italics indicates the mismatch introduced into safety BD to mask the BP sequence in the PTM.

       A double *trans*-splicing PTM construct (DS-PTM) was also made adding a 5' splice site and a second target binding domain complementary to the second intron of βHCG pre-mRNA to the 3' end of the toxin coding sequence of PTM+SF (Figure A).

6.1.2. βHCG6 TARGET PRE-mRNA

To produce the *in vitro* target pre-mRNA, a SacI fragment of βHCG gene 6 (accession #X00266) was cloned into pBS(-). This produced an 805 bp insert from nucleotide 460 to 1265, which includes the 5' untranslated region, initiation

5 codon, exon 1, intron 1, exon 2, and most of intron 2. For *in vivo* studies, an EcoRI and BamHI fragment was cloned into mammalian expression vector (pc3.1DNA), producing βHCG6.

6.1.3. mRNA PREPARATION

For *in vitro* splicing experiments, βHCG6, β-globin pre-mRNA and

10 different PTM mRNAs were synthesized by *in vitro* transcription of BamHI and HindIII digested plasmid DNAs respectively, using T7 mRNA polymerase (Pasman & Garcia-Blanco, 1996, Nucleic Acids Res. 24:1638). Synthesized mRNAs were purified by electrophoresis on a denaturing polyacrylamide gel, and the products were excised and eluted.

15 ## 6.1.4 *IN VITRO* SPLICING

PTMs and target pre-mRNA were annealed by heating at 98°C followed by slow cooling to 30-34°C. Each reaction contained 4 μl of annealed mRNA complex (100 ng of target and 200 ng of PTM), 1X splice buffer (2 Memorandum $MgCl_2$, 1 Memorandum ATP, 5 Memorandum creatinine phosphate,

20 and 40 Memorandum KCI) and 4 μl of HeLa splice nuclear extract (Promega) in a 12.5 μl final volume. Reactions were incubated at 30°C for the indicated times and stopped by the addition of an equal volume of high salt buffer (7 M urea, 5% SDS, 100 Memorandum LiCl, 10 Memorandum EDTA and 10 Memorandum TrisHCI, pH 7.5). Nucleic acids were purified by extraction with phenol:chloroform:isoamyl

25 alcohol (50:49:1) followed by ethanol precipitation.

6.1.5. REVERSE TRANSCRIPTION-PCR REACTIONS

RT-PCR analysis was performed using EZ-RT PCR kit (Perkin-Elmer, Foster City, CA). Each reaction contained 10 ng of *cis*- or *trans*-spliced mRNA, or 1-2 μg of total mRNA, 0.1 μl of each 3' and 5' specific primer, 0.3 Memorandum of

32

each dNTP, 1X EZ buffer (50 Memorandum bicine, 115 Memorandum potassium

acetate, 4% glycerol, pH 8.2), 2.5 Memorandum magnesium acetate and 5 U of r*Tth*

DNA polymerase in a 50 μl reaction volume. Reverse transcription was performed at

60°C for 45 min followed by PCR amplification of the resulting cDNA as follows:

5   one cycle of initial denaturation at 94°C for 30 sec, and 25 cycles of denaturation at

94°C for 18 sec and annealing and extension at 60°C for 40 sec, followed by a 7 min

final extension at 70°C. Reaction products were separated by electrophoresis in

agarose gels.

Primers used in the study were as follows:

10   DT-1F: GGCGCTGCAGGGCGCTGATGATGTTGTTG

DT-2R: GGCGAAGCTTGGATCCGACACGATTTCCTGCACAGG

DT-3R: CATCGTCATAATTTCCTTGTG

DT-4R: ATGGAATCTACATAACCAGG

DT-5R: GAAGGCTGAGCACTACACGC

15   HCG-R2: CGGCACCGTGGCCGAAGTGG,

Bio-HCG-F: ACCGGAATTCATGAAGCCAGGTACACCAGG

β-globulin-F: GGGCAAGGTGAACGTGGATG

β-globulin-R: ATCAGGAGTGGACAGATCC

6.1.6 CELL GROWTH, TRANSFECTION AND mRNA

20   ISOLATION

Human lung cancer cell line H1299 (ATCC accession # CRL-5803)

was grown in RPMI medium supplemented with 10% fetal bovine serum at 37°C in a

5% CO_2 environment. Cells were transfected with pcSp+CRM (CRM is a non-

functional toxin), a vector expressing a PTM, or vector alone (pcDNA3.1) using

25   lipofectamine reagent (Life Technologies, Gaithersburg, MD). The assay was scored

for neomycin resistance (neo[r]) colony formation two weeks after transfection. Four

neo[r] colonies were selected and expanded under continued neo selection. Total

cellular mRNA was isolated using RNA exol (BioChain Institute, Inc., San Leandro,

CA) and used for RT-PCR.

33

6.1.7. *TRANS*-SPLICING IN TUMORS IN NUDE MICE

Eleven nude mice were bilaterally injected (except B10, B11 and B12 had 1 tumor) into the dorsal flank subcutaneous space with 1 x 10^7 H1299 human lung tumor cells (day 1). On day 14, the mice were given an appropriate dose of anesthesia

5   and injected with, or without electroporation (T820, BTX Inc., San Diego, CA) in several orientations with a total volume of 100 μl of saline containing 100 μg pcSp+CRM with or without pcβHCG6 or pcPTM+Sp. Solutions injected into the right side tumors also contained India ink to mark needle tracks. The animals were sacrificed 48 hours later and the tumor excised and immediately frozen at -80°C. For

10   analysis, 10 mg of each tumor was homogenized and mRNA was isolated using a Dynabeads mRNA direct kit (Dynal) following the manufacturers directions. Purified mRNA (2 μl of 10 μl total volume) was subjected to RT-PCR using βHCG-F and DT-5R primers as described earlier. All samples were re-amplified using DT-3R, a nested DT-A primer and biotinylated βHCG-F and the products were analyzed by

15   electrophoresis on a 2% agarose gel. Samples that produced a band were processed into single stranded DNA using M280 Streptavidin Dynabeads and sequenced using a toxin specific primer (DT-3R).

6.2. RESULTS

6.2.1. SYNTHESIS OF PTM

20   A prototypical *trans*-splicing mRNA molecule, pcPTM+Sp (Figure 1A) was constructed that included: an 18 nt target binding domain (complementary to βHCG6 intron 1), a 30 nucleotide spacer region, branch point (BP) sequence, a polypyrimidine tract (PPT) and an AG dinucleotide at the 3' splice site immediately upstream of an exon encoding diphtheria toxin subunit A (DT-A) (Uchida *et al.*, 1973,

25   J. Biol. Chem. 248:3838). Later DT-A exons were modified to eliminate translation initiation sites at codon 14. The PTM constructs were designed for maximal activity in order to demonstrate *trans*-splicing; therefore, they included potent 3' splice elements (yeast BP and a mammalian PPT) (Moore *et al.*, 1993, In The mRNA World, R.F. Gesteland and J.F. Atkins, eds. (Cold Spring Harbor, New York: Cold Spring

30   Harbor Laboratory Press). βHCG6 pre-mRNA (Talmadge *et al.*, 1984, Nucleic Acids

34

Res. 12:8415) was chosen as a model target as this gene is expressed in most tumor

cells. It is not expressed in normal adult cells, with the exception of some in the

pituitary gland and gonads. (Acevedo *et al.*, 1992, Cancer 76:1467; Hoon *et al.*, 1996,

Int J. Cancer 69:369; Bellet *et al.*, 1997, Cancer Res. 57:516). As shown in Figure

5   1C, pcPTM+Sp forms conventional Watson-Crick base pairs by its binding domain

with the 3' end of βHCG6 intron 1, masking the intronic 3' splice signals of the target.

This feature is designed to facilitate *trans*-splicing between the target and the PTM.

HeLa nuclear extracts were used in conjunction with established

splicing procedures (Pasman & Garcia-Blanco, 1996, Nucleic Acids Res. 24:1638) to

10   test if a PTM construct could invade the βHCG6 pre-mRNA target. The products of

in vitro trans-splicing were detected by RT-PCR, using primers specific for chimeric

mRNA molecules. The predicted product of a successful *trans*-splicing reaction is a

chimeric mRNA comprising the first exon of βHCG6, followed immediately by the

exon contributed from pcPTM+Sp encoding DT-A (Figure 1C). Such chimeric

15   mRNAs were readily detected by RT-PCR using primers βHCG-F (specific to

βHCG6 exon 1) and DT-3R (specific to DT-A, Figure 2, lanes 1-2). At time zero or

in the absence of ATP, no 466 bp product was observed, indicating that this reaction

was both ATP and time dependent.

The target binding domain of pcPTM+Sp contained 18 nucleotides

20   complementary to βHCG6 intron 1 pre-mRNA and demonstrated efficient *trans*-

splicing (Figure 2A, lanes 1-2). *Trans*-splicing efficiency decreased at least 8 fold

(Figure 2, lanes 3-4) using non-targeted PTM-Sp, which contains a non-

complementary 18 nucleotide "non-binding domain". *Trans*-splicing efficiencies of

PTM mRNAs with or without a spacer between the binding domain and BP were also

25   compared. This experiment demonstrated a significant increase in the efficiency of

trans-splicing by the addition of a spacer (Figure 2B, lanes 2 + 5). To facilitate the

recruitment of splicing factors required for efficient *trans*-splicing, some space may be

needed between the 3' splice site and the double stranded secondary structure

produced by the binding domain/target interaction.

30   To investigate the effect of PTM length on *trans*-splicing specificity,

shorter PTMs were synthesized from AccI cut PTM plasmid (see Figure 1). This

35

eliminated 479 nt from the 3' end of the DT-A coding sequence. Figure 2B shows the trans-splicing ability of a targeted short PTM(+) (lanes 10-12), compared to a non-targeted short PTM(-) (lanes 14-17). Short PTM+ produced substantially more *trans-*spliced product (Figure 2B, lane 12) than its counterpart, non-targeted short PTM

5   (Figure 2B, lane 17). These experiments indicate that longer PTMs may have increased potential to mediate *trans*-splicing non-specifically.

6.2.2.   ACCURACY OF PTM SPLICEOSOME MEDIATED TRANS
SPLICING

To confirm that *trans*-splicing between the pcPTM+Sp and βHCG6

10   target is precise, RT-PCR amplified product was produced using 5' biotinylated βHCG-F and nonbiotinylated DT-3R primers. This product was converted into single stranded DNA and sequenced directly with primer DT-3R (DT-A specific reverse primer) using the method of Mitchell and Merril (1989, Anal. Biochem. 178:239). *Trans*-splicing occurred exactly between the predicted splice sites (Figure 3),

15   confirming that a conventional pre-mRNA can be invaded by an engineered PTM construct during splicing; moreover, this reaction is precise.

In addition selective *trans*-splicing of a double splicing PTM (DS-PTM) was observed (Figure 8B). The DS-PTM can produce *trans*-splicing by contributing either a 3' or 5' splice site. Further, DS-PTMs can be constructed which

20   will be capable of simultaneously double trans-splicing, at both a 3' and 5' site, thereby permitting exon replacement. Figure 8B demonstrates that in this construct the 5' splice site is most active at a 1:1 concentration of target βHCG pre-mRNA:DS-PTM. At a 1:6 ratio the 3' splice site is more active.

6.2.3.   3' SPLICE SITES ARE ESSENTIAL FOR PTM *TRANS*-SPLICING

25   In general, the 3' splice site contains three elements: 1) a BP sequence located 5' of the acceptor site, 2) a PPT consisting of a short run of pyrimidine residues, and 3) a YAG trinucleotide splice site acceptor at the intron-exon border (Senapathy *et al.*, 1990, Cell 91:875; Moore *et al.*, 1993). Deletion or alteration of

36

one of these sequence elements are known to either decrease or abolish splicing (Aebi *et al.*, 1986; Reed & Maniatis 1988, Genes Dev. 2:1268; Reed, 1989, Genes Dev. 3:2113; Roscigno *et al.*, 1993, J. Biol. Chem. 268:11222; Coolidge *et al.*, 1997, Nucleic Acids Res. 25:888). The role of these conserved elements in targeted *trans*-

5   splicing was addressed experimentally. In one case [(BP(-)Py(-)AG(-)], all three *cis.*-elements (BP, PPT and AG dinucleotide) were replaced by random sequences. A second splicing mutant [(Py(-)AG(-)] was constructed in which the PPT and the 3' splice site acceptor were mutated and substituted by random sequences. Neither construct was able to support *trans*-splicing *in vitro* (Figure 2A, lanes 5-8), suggesting

10   that, as in the case of conventional *cis*-splicing, the PTM *trans*-splicing process also requires a functional BP, PPT and AG acceptor at the 3' splice site.

6.2.4.   DEVELOPMENT OF A "SAFETY" SPLICE SITE
TO INCREASE SPECIFICITY

To improve the levels of target specificity achieved by the inclusion of

15   a binding domain or by shortening the PTM, the target-binding domain of several PTM constructs was modified to create an intra-molecular stem to mask the 3' splice site (termed a "safety PTM"). The safety stem is formed by portions of the binding domain that partially base pair with regions of the PTM 3' splice site or sequences adjacent to them, thereby blocking the access of spliceosomal components to the PTM

20   3' splice site prior to target acquisition (Figure 4A, PTM+SF). Base pairing between free portions of the PTM binding domain and βHCG6 target region unwinds the safety stem, allowing splicing factors such as U2AF to bind to the PTM 3' splice site and initiate *trans*-splicing (Figure 4B).

This concept was tested in splicing reactions containing either

25   PTM+SF (safety) or pcPTM+Sp (linear), and both target (βHCG6) and non-target (β-globin) pre-mRNA. The spliced products were subsequently analyzed by RT-PCR and gel electrophoresis. Using βHCG-F and DT-3R primers, the specific 196 bp *trans*-spliced band was demonstrated in reactions containing βHCG target and either linear PTM (pcPTM+Sp, Figure 5, lane 2) or safety PTM (PTM+SF, Figure 5, lane 8).

30   Comparison of the targeted *trans*-splicing between linear PTM (Figure 5, lane 2) and

safety PTM (Figure 5, lane 8) demonstrated that the safety PTM *trans*-spliced less efficiently than the linear PTM.

Non-targeted reactions were amplified using β-globin-F (specific to exon 1 of β-globin) and DT-3R primers. The predicted product generated by non-
5   specific PTM *trans*-splicing with β-globin pre-mRNA is 189 bp. Non-specific *trans*-splicing was evident between linear PTM and β-globin pre-mRNA (Figure 5, lane 5). In contrast, non-specific *trans*-splicing was virtually eliminated by the use of safety PTM (Figure 5, lane 11). This was not unexpected, since the linear PTM was designed for maximal activity to prove the concept of spliceosome-mediated *trans*-
10   splicing. The open structure of the linear PTM combined with its potent 3' splice sites strongly promotes the binding of splicing factors. Once bound, these splicing factors can potentially initiate *trans*-splicing with any 5' splice site, in a process similar to *trans*-splicing in trypanosomes. The safety stem was designed to prevent splicing factors, such as U2AF from binding to the PTM prior to target acquisition. This result
15   is consistent with a model that base-pairing between the free portion of the binding domain and the βHCG6 target unwinds the safety stem (by mRNA-mRNA interaction), uncovering the 3' splice site, permitting the recruitment of splicing factors and initiation of *trans*-splicing. No *trans*-splicing was detected between β-globin and βHCG6 pre-mRNAs (Figure 5, lanes 3, 6, 9 and 12).

20   6.2.5   *IN VITRO TRANS*-SPLICING OF SAFETY PTM AND VARIANTS

To better understand the role of *cis*-elements at the 3' splice site in *trans*-splicing a series of safety PTM variants were constructed in which either the PPT was weakened by substitution with purines and/or the BP was modified by base substitution (see Table I). *In vitro trans*-splicing efficiency of the safety (PTM+SF)
25   was compared to three safety variants, which demonstrated a decreased ability to *trans*-splice. The greatest effect was observed with varient 2 (PTM+SFPy2), which was *trans*-splicing incompetent (Figure 4C, lanes 5-6). This inhibition of *trans*-splicing may be attributed to a weakened PPT and/or the higher T_m of the safety stem. In contrast, variations in the BP sequence (PTM+SFBP3) did not markedly effect
30   *trans*-splicing (Figure 4C, lanes 7-8). This was not surprising since the modifications

38

introduced were within the mammalian branch point consensus range YNYURAC

(where Y = pyrimidine, R = purine and N = any nucleotide) (Moore *et al.*, 1993).

This finding indicates that the branch point sequence can be removed without

affecting splicing efficiency. Alterations in the PPT (PTM+SF-Py1) decreased the

5   level of *trans*-splicing (lanes 3-4). Similarly, when both BP and PPT were altered

PTM+SFBP3-Py1, they caused a further reduction in *trans*-splicing (Figure 4C, lanes

9-10). The order of *trans*-splicing efficiency of these safety variants is

PTM+SF>PTM+SFBP3> PTM+SFPy1>PTM+SFBP3-Py1>PTM+SFPy2. These

results confirm that both the PPT and BP are important for efficient *in vitro trans*-

10   splicing (Roscigno *et al.*, 1993, J. Biol. Chem. 268:11222).

6.2.6   COMPETITION BETWEEN *CIS*- AND *TRANS*- SPLICING

To determine if it was possible to block pre-mRNA *cis*-splicing by

increasing concentrations of PTM, experiments were performed to drive the reaction

towards *trans*-splicing. Splicing reactions were conducted with a constant amount of

15   βHCG6 pre-mRNA target and various concentrations of *trans*-splicing PTM. *Cis*-

splicing was monitored by RT-PCR using primers to βHCG-F (exon 1) and βHCG-R2

(exon 2). This amplified the expected 125 bp *cis*-spliced and 478 bp unspliced

products (Figure 6A). The primers βHCG-F and DT-3R were used to detect *trans*-

spliced products (Figure 6B). At lower concentrations of PTM, *cis*-splicing (Fig. 6A,

20   lanes 1-4) predominated over *trans*-splicing (Figure 6B, lanes 1-4). *Cis*-splicing was

reduced approximately by 50% at a PTM concentration 1.5 fold greater than target.

Increasing the PTM mRNA concentration to 3 fold that of target inhibited *cis*-splicing

by more than 90% (Figure 6A, lanes 7-9), with a concomitant increase in the *trans*-

spliced product (Figure 6B, lanes 6-10). A competitive RT-PCR was performed to

25   simultaneously amplify both *cis* and *trans*-spliced products by including all three

primers (βHCG-F, HCG-R2 and DT-3R) in a single reaction. This experiment had

similar results to those seen in Figure 6, demonstrating that under *in vitro* conditions,

a PTM can effectively block target pre-mRNA *cis*-splicing and replace it with the

production of an engineered *trans*-spliced chimeric mRNA.

39

6.2.7. _TRANS_-SPLICING IN TISSUE CULTURE

To demonstrate the mechanism of _trans_-splicing in a cell culture model, the human lung cancer line H1299 (βHCG6 positive) was transfected with a vector expressing SP+CRM (a non-functional diphtheria toxin) or vector alone (pcDNA3.1) and grown in the presence of neomycin. Four neomycin resistant colonies were individually collected after 14 days and expanded in the continued presence of neomycin. Total mRNA was isolated from each clone and analyzed by RT-PCR using primers βHCG-F and DT-3R. This yielded the predicted 196 bp _trans_-spliced product in three out of the four selected clones (Figure 7A, lanes 2, 3 and 4). The amplified product from clone #2 was directly sequenced, confirming that PTM driven _trans_-splicing occurred in human cells exactly at the predicted splice sites of endogenously expressed βHCG6 target exon 1 and the first nucleotide of DT-A (Figure 7B).

6.2.8. _TRANS_-SPLICING IN AN _IN VIVO_ MODEL

To demonstrate the mechanism of _trans_-splicing _in vivo_, the following experiment was conducted in athymic (nude) mice. Tumors were established by injecting 10^7 H1299 cells into the dorsal flank subcutaneous space. On day 14, PTM expression plasmids were injected into tumors. Most tumors were then subjected to electroporation to facilitate plasmid delivery (see Table 2, below). After 48 hrs, tumors were removed, poly-A mRNA was isolated and amplified by RT-PCR. _Trans_-splicing was detected in 8 out of 19 PTM treated tumors. Two samples produced the predicted _trans_-spliced product (466 bp) from mRNA after one round of RT-PCR. Six additional tumors were subsequently positive for _trans_-splicing by a second PCR amplification using a nested set of primers that produced the predicted 196 bp product (Table 2). Each positive sample was sequenced, demonstrating that βHCG6 exon 1 was precisely _trans_-spliced to the coding sequence of DT-A (wild type or CRM mutant) at the predicted splice sites. Six of the positive samples were from treatment groups that received cotransfected plasmids, pcPTM+CRM and pcHCG6, which artificially increased the concentration of target pre-mRNA. This was done to enhance the probability of detecting _trans_-spliced events. The other two positive

tumors were from a group that received only pcPTM+Sp (wild type DT-A). These tumors were not transfected with βHCG6 expression plasmid, demonstrating once again, as in the tissue culture model described in Section 6.2.7, that *trans*-splicing occurred between the PTM and endogenous βHCG6 pre-mRNA produced by tumor

5   cells.

41

Table 2. *Trans*-splicing in tumors in nude mice.							
Group	Plasmid	Tumor		Electroporation	RT-PCR (For 466 bp)		Nested PCR (For 196 bp)	Splice junction sequence
		Left	Right		Left	Right		
[1]B1	pCMV-Sport	B1-1	B1-2	-	-	-	-	-	-
B2	pCMV-Sport	B1-3	B1-4	[a]1000V/cm	-	-	-	-	-
B3	pcSp+CRM	B3-1	B3-2	[a]1000V/cm	-	-	-	-	-
		B3-3	B3-4	[a]1000V/cm	-	-	-	-	-
B4	pcSp+CRM	B4-1	B4-2	[b]50V/cm	-	-	-	-	-
		B4-3	B4-4	[c]25V/cm	-	-	-	-	-
B5	pcSp+CRM/ pcHCG6	B5-1	B5-2	[a]1000V/cm	+	-	+	+	ATGTTCCAGι GGCGTGATG AT
		B5-3	B5-4	[a]1000V/cm	+	-	+	+	ATGTTCCAGι GGCGTGATG AT
B6	pcSp+CRM/ pcHCG6	B6-1	B6-2	[b]50V/cm	-	-	-	-	-
		B6-3	B6-4	[c]25V/cm	-	-	+	+	ATGTTCCAGι GGCGTGATG AT
B7	pc PTM+Sp	B7-1		[a]1000V/cm	-		-		-
B8	pc PTM+Sp	B8-1		[b]50V/cm	-		+		ATGTTCCAGι GGCGTGATG AT
[1]B9	pc PTM+Sp	B9-1		-	-		+		ATGTTCCAGι GGCGTGATG AT

[a]: 6 pulses of 99μs sets of 3 pulses administered orthogonally

[b]: 8 pulses of 10ms sets of 4 pulses administered orthogonally

42

^c: 8 pulses of 50ms sets of 4 pulses administered orthogonally

¹: did not receive electroporation

7.   EXAMPLE: _lacZ TRANS-_SPLICING MODEL

In order to demonstrate and evaluate the generality of the mechanism
of spliceosome mediated targeted _trans_-splicing between a specific pre-mRNA target
and a PTM, a simple model system based on expression of enzyme β-galactosidase
was developed. The following section describes results demonstrating successful
splicesome mediated targeted _trans_-splicing between a specific target and a PTM.

7.1   MATERIALS AND METHODS

7.1.1.   PRIMER SEQUENCES

The following primers were used for testing the _lacZ_ model system:

5' Lac-1F

GCATGAATTCGGTACCATGGGGGGGGTTCTCATCATCATC

5' Lac-1R   CTGAGGATCCTCTTACCTGTAAACGCCCATACTGAC

3' Lac-1F

GCATGGTAACCCTGCAGGGCGGCTTCGTCTGGGACTGG

3' Lac-1R   CTGAAAGCTTGTTAACTTATTATTTTTGACACCAGACC

3' Lac-Stop

GCATGGTAACCCTGCAGGGCGGCTTCGTCTAATAATGGGACTGGGTG

HCG-In1F   GCATGGATCCTCCGGAGGGCCCCTGGGCACCTTCCAC

HCG-In1R

CTGACTGCAGGGTAACCGGACAAGGACACTGCTTCACC

HCG-Ex2F   GCATGGTAACCCTGCAGGGGCTGCTGCTGTTGCTG

HCG-Ex2R   CTGAAAGCTTGTTAACCAGCTCACCATGGTGGGGCAG

Lac-TR1 (Biotin): 7-GGCTTTCGCTACCTGGAGAGAC

Lac-TR2   GCTGGATGCGGCGTGCGGTCG

HCG-R2:   CGGCACCGTGGCCGAAGTGG

43

7. 1.2. CONSTRUCTION OF THE *lacZ* PRE-mRNA TARGET MOLECULE

The lacZ target 1 pre-mRNA (pc3.1 lacT1) was constructed by cloning of the following three PCR products: (i) the 5' fragment of lacZ; followed by (ii) βHCG6 intron 1; (iii) and the 3' fragment of lacZ. The 5' and 3' fragment of the lacZ gene were PCR amplified from template pcDNA3.1/His/LacZ (Invitrogen,San Diego, CA) using the following primers: 5' Lac-1F and 5'Lac-1R (for 5' fragment), and 3'Lac-1F and 3' Lac-1R (for 3' fragment). The amplified lacZ 5' fragment is 1788 bp long which includes the initiation codon, and the amplified 3' fragment is 1385 bp long and has the natural 5' and 3' splice sites in addition to a branch point, polypyrimidine tract and βHCG6 intron 1. The βHCG6 intron 1 was PCR amplified using the following primers: HCG-In1F and HCG-In1R.

The lacZ target 2 is an identical version of lacZ target 1 except it contains two stop codons (TAA TAA) in frame four codons after the 3'splice site. This was created by PCR amplification of the 3' fragment (lacZ) using the following primers: 3' Lac-Stop and 3' Lac 1R and replacing the functional 3' fragment in LacZ target 1.

7.1.3. CONSTRUCTION OF pc3.1 PTM1 and pc3.1 PTM2

The pre-trans-splicing molecule, pc3.1 PTM1 was created by digesting pPTM +Sp with PstI and HindIII and replacing the DNA fragment encoding the DT-A toxin with the a DNA fragment encoding the functional 3' end of lacZ. This fragment was generated by PCR amplification using the following primers: 3' Lac-1F and 3' Lac-1R. For cell culture experiments, an EcoRI and HindIII fragment of pc3.1 PTM2 which contains the binding domain to HCG intron 1, a 30 bp spacer, a yeast branch point (TACTAAC), and strong polypyrimidine tract followed by the lacZ cloned was cloned into pcDNA3.1.

The pre-trans-splicing molecule, pc3.1 PTM2 was created by digesting pPTM +Sp with PstI and HindIII and replacing the DNA fragment encoding the DT-A toxin with the βHCG6 exon 2. βHCG6 exon 2 was generated by PCR amplification using the following primers: HCG-Ex2F and HCG-Ex2R. For cell culture

44

experiments, an EcoRI and HindIII fragment of pc3.1 PTM2 which contains the

binding domain to HCG intron 1, a 30 bp spacer, a yeast branch point (TACTAAC),

and strong polypyrimidine tract followed by the βHCG6 exon 2 cloned was used.

7.1.4. CO-TRANSFECTION OF THE *lacZ* SPLICE
TARGET PRE-mRNA AND PTMS INTO 293T
<u>CELLS</u>

Human embryonic kidney cells (293T) were grown in DMEM medium

supplemented with 10% FBS at 37°C in a 5% CO_2. Cells were co-transfected with

pc3.1 LacT1 and pc3.1 PTM2, or pc3.1 LacT2 and pc3.1 PTM, using Lipofectamine

Plus (Life Technologies,Gaithersburg, MD) according to the manufacturer's

instructions. 24 hours post-transfection, the cells were harvested; total RNA was

isolated and RT-PCR was performed using specific primers for the target and PTM

molecules. β-galactosidase activity was also monitored by staining the cells using a

β-gal staining kit (Invitrogen, San Deigo. CA).

7.2. <u>RESULTS</u>

7.2.1 THE *lacZ* SPLICE TARGET *CIS*-SPLICES EFFICIENTLY
<u>TO PRODUCE FUNCTIONAL β-GALACTOSIDASE</u>

To test the ability of the splice target pre-mRNA to *cis*-splice

efficiently, pc3.1 *lac*T1 was transfected into 293 T cells using Lipfectamine Plus

reagent (Life Technologies,Gaithersburg, MD) followed by RT-PCR analysis of total

RNA. Sequence analysis of the *cis*-spliced RT-PCR product indicated that splicing

was accurate and occurred exactly at the predicted splice sites (Fig. 12B). In addition,

accurate cis-splicing of the target pre-mRNA molecule results in formation of a

mRNA capable of encoding active β-galactosidase which catalyzes the hydrolysis of

β-galactosidase, *i.e.*, X-gal, producing a blue color that can be visualized under a

microscope. Accurate *cis*-splicing of the target pre-mRNA was further confirmed by

successfully detecting β-galactosidase enzyme activity.

45

Repair of defective lacZ target 2 pre-mRNA by trans-splicing of the functional 3' lacZ fragment (PTM1) was measured by staining for β-galactosidase enzyme activity. For this purpose, 293T cells were co-transfected with lacZ target 2 pre-mRNA (containing a defective 3' fragment) and PTM1 (containing functional 3'

5   lacZ activity). 24 hours post-transfection cells were assayed for β-galactosidase enzyme activity. Efficient trans-splicing of PTM1 into the lacZ target 2 pre-mRNA will result in the production of functional β-galactosidase activity. As demonstrated in Figure 11B-E, trans-splicing of PTM 1 into lacZ target 2 results in restoration of β-galactosidase enzyme activity up to 5% to 10% compared to control.

10   7.2.2.   TARGETED *TRANS*-SPLICING BETWEEN

                                THE lacZ TARGET PRE-mRNA and PTM2.

To assay for *trans*-splicing, *lacZ* target pre-mRNA and PTM2 were transfected into 293 T cells. Following transfection, total RNA was analyzed using RT-PCR. The following primers were used in the PCR reactions: *lacZ*-TR1 (lacZ 5'

15   exon specific) and HCGR2 (βHCGR exon 2 specific). The RT PCR reaction produced the expected 195 bp *trans*-spliced product (Fig. 11, lanes 2 and 3) demonstrating efficient *trans*-splicing between the *lacZ* target pre-mRNA and PTM 2. Lane 1 represents the control, which does not contain PTM 2.

The efficiency of the *trans*-splicing was also measured by staining for

20   β-galactosidase enzyme activity. To assay for trans-splicing, 293T cells were co-transfected with lacZ target pre-mRNA and PTM 2. 24 hours post-transfection, cells were assayed for β-galactosidase activity. If there is efficient *trans*-splicing between the target pre-mRNA and the PTM, a chimeric mRNA is produced consisting of the 5' fragment of the lacZ target pre-mRNA and βHCG6 exon 2 is formed which is

25   incapable of coding for an active β-galactosidase. Results from the co-transfection experiments demonstrated that *trans*-splicing of PTM2 into lacZ target 1 resulted in the reduction of β-galactosidase activity by 10-15% compared to the control.

To further confirm that *trans*-splicing between the lacZ target pre-mRNA and PTM2 is accurate, RT-PCR was performed using 5' biotinylated lacZ-

30   TR1 and non-biotinylated HCGR2 primers. Single stranded DNA was isolated and sequenced directly using HCGR2 primer (HCG exon 2 specific primer). As

46

evidenced by the sequence of the splice junction, *trans*-splicing occurred exactly as predicted between the splice sites (Fig. 12A and 12B), confirming that a conventional pre-mRNA can be invaded by an engineered PTM during splicing, and moreover, that this reaction is precise.

5   8.   EXAMPLE: CORRECTION OF THE CYSTIC FIBROSIS
                           TRANSMEMBRANE REGUALTOR GENE

         Cystic fibrosis (CF) is one of the most common genetic diseases in the world. The gene associated with CF has been isolated and its protein product deduced (Kerem, B.S. et al., 1989, Science 245:1073-1080; Riordan et al., 1989,

10   Science 245:1066-1073;Rommans, et al., 1989, Science 245:1059-1065). The protein product of the CF associated gene is referred to as the cystic fibrosis trans-membrane conductance regulator (CFTR). The most common disease-causing mutation which accounts for ~70% of all mutant alleles is a deletion of three nucleotides in exon 10 that encode for a phenylalanine at position 508 (ΔF508). The following section

15   describes the successful repair of the cystic fibrosis gene using spliceosome mediated *trans*-splicing and demonstrates the feasibility of repairing CFTR in a model system.

                    8.1   MATERIALS AND METHODS
                    8.1.1   PRE-TRANS-SPLICING MOLECULE

20   The CFTR pre-trans-splicing molecule (PTM) consists of a 23 nucleotide binding domain complimentary to CFTR intron 9 (3' end, -13 to -31), a 30 nucleotide spacer region (to allow efficient binding of spliceosomal components), branch point (BP) sequence, polypyrimidine tract (PPT) and an AG dinucleotide at the 3' splice site immediately upstream of the sequence encoding CFTR exon 10 (wild

25   type sequence containing F508). This initial PTM was designed for maximal activity in order to demonstrate *trans*-splicing; therefore the PTM included a UACUAAC yeast consensus BP sequence and an extensive PPT. An 18 nucleotide HIS tag (6 histamine codons) was included after wild type exon 10 coding sequence to allow specific amplification and isolation of the *trans*-spliced products and not the

30   endogenous CFTR. The oligonucleotides used to generate the two fragments

47

included unique restriction sites. (ApaI and PstI, and PstI and NotI, respectively) to facilitate directed cloning of amplified DNA into the mammalian expression vector pcDNA3.1.

8.1.2. THE TARGET CFTR PRE-mRNA MINI-GENE

The CFTR mini-gene target is shown in Figure 13 and consists of CFTR exon 9 ; the functional 5' and 3' regions of intron 9 (260 and 265 nucleotides from each end, respectively); exon 10 [ΔF508]; and the 5' region of intron 10 (96 nucleotides). In addition, as depicted in Figure 16, a mini-target gene comprising CFTR exons 1-9 and 10-24 can be used to test the use of spliceosome mediated *trans*-splicing for correction of the cystic fibrosis mutation. Figure 17, shows a double splicing PTM that may also be used for correction of the cystic fibrosis mutation. As shown, the double splicing PTM contains CFTR BD intron 9, a spacer, a branch point, a polypyrimidine tract, a 3' splice site, CFTR exon 10, a spacer, a branch point, a polypyrimidine tract, a 5' splice site and CFTR BD exon 10.

8.1.3 OLIGONUCLEOTIDES

The following oligonucleotides were used to create CFTR PTM:

Forward   CF3

      ACCT   GGGCCC ACC CAT TAT TAG GTC ATT AT *CCGCGG* AAC ATT ATA

      Apal site.   Intron 9 CFTR, -12 to -34.

Reverse   CF4

      ACCT   CTGCA*GGTGACC* CTG CAG GAA AAA AAA GAA G

      PstI.   BstEI.   PPT.

48

<u>Forward</u>   <u>CF5</u>

ACCT   CTGCAG   ACT TCA CTT CTA ATG ATG AT

        PstI.   Exon 10 CFTR, +1 to +24

<u>Reverse</u>.   <u>CF6</u>

5   ACCT   GCGGCCGC CTA ATG ATG ATG ATG ATG CTC TTC TAG
TTG GCA TGC

        Not I.   Stop Polyhistamine tag   Exon 10 CFTR, +15 to +132

The following nucleotides were used to create the CFTR TARGET
pre-mRNA mini gene (Exon 9 + mini-Intron 9 + Exon 10 + 5' end Intron 10):

10   <u>Forward</u>   <u>CF18</u>

GACCT   CTCGAG GGA TTT GGG GAA TTA TTT GAG

        XhoI   Exon 9 CFTR, 1 to 21.

<u>Reverse</u>   <u>CF19</u>

CTGACCT GCGGCCGC TAC AGT GTT GAA TGT GGT GC

15   NotI.   Intron 9 5' end.

<u>Forward</u>   <u>CF20</u>

CTGACCT GCGGCCGC CCA ACT ATC TGA ATC ATG TG

        NotI.   Intron 9 3' end.

<u>Reverse</u>   <u>CF21</u>

20   GACCT   CTTAAG TAG ACT AAC CGA TTG AAT ATG

        AflII   Intron 10 5' end.

The following oligonucleotides were used for detection of trans-
spliced products:

<u>Reverse</u>   <u>Bio-His</u>

25   CTA ATG ATG ATG ATG ATG ATG

    Stop. Polyhistidine tag (5' biotin label).

Reverse   Bio-His(2)

                 CGC CTA ATG ATG ATG ATG ATG

                 3' UT Stop. Polyhistidine tag (5' biotin label).

Forward   CF8

5   CTT CTT GGT ACT CCT GTC CTG

                 Exon 9 CFTR.

Forward   CF18

                 GACCT CTCGAG GGA TTT GGG GAA TTA TTT GAG

                          Xhol.   Exon 9 CFTR.

10   Reverse CF28

                 AAC TAG AAG GCA CAG TCG AGG

                 Pc3.1 vector sequence (present in PTM 3' UT but not target).

8.2.   RESULTS

The PTM and target pre-mRNA were co-transfected in 293 embryonic

15   kidney cells using lipofectamine (Life Technologies,Gaithersburg, MD). Cells were

harvested 24 h post transfection and RNA was isolated. Using PTM and target-

specific primers in RT-PCR reactions, a *trans*-spliced product was detected in which

mutant exon 10 of the target pre-mRNA was replaced by the wild type exon 10 of the

PTM (Figure14). Sequence analysis of the *trans*-spliced product confirmed the

20   restoration of the three nucleotide deletion and that splicing was accurate, occurring at

the predicted splice sites (Figure 15), demonstrating for the first time RNA repair of

the cystic fibrosis gene, CFTR.

The present invention is not to be limited in scope by the specific

embodiments described herein. Indeed, various modifications of the invention in

25   addition to those described herein will become apparent to those skilled in the art

from the foregoing description and accompanying Figures. Such modifications are

intended to fall within the scope of the appended claims. Various references are cited herein, the disclosure of which are incorporated by reference in their entireties.

51

CLAIMS

1. A pre-*trans*-splicing molecule comprising:

   a)   one or more target binding domains that target binding
        of the pre-*trans*-splicing molecule to a target pre-
        mRNA;

   b)   a 3' splice region comprising a branch point, a
        pyrimidine tract and a 3' splice acceptor site;

   c)   a spacer region that separates the mRNA splice region
        from the target binding domain; and

   d)   a nucleotide sequence to be trans-spliced to the target
        pre-mRNA.

2. A pre-*trans*-splicing molecule comprising:

   a)   one or more target binding domains that target binding
        of the pre-*trans*-splicing molecule to a target pre-
        mRNA;

   b)   a 3' splice region comprising a branch point, a
        pyrimidine tract and a 3' splice acceptor site and or a 5'
        splice donor site;

   c)   a spacer region that separates the splice region from the
        target binding domain; and

   d)   a nucleotide sequence to be trans-spliced to the target
        pre-mRNA.

3. The molecule of claim 1 or 2 further comprising nucleotide
   sequences encoding the cystic fibrosis trans-membrane
   regulator.

4. The molecule of claim 1 or 2 further comprising exon 10 cystic
   fibrosis trans-membrane regulator nucleotide sequences.

52

5.   An isolated nucleic acid molecule comprising a nucleotide
        sequence encoding a pre-*trans*-splicing molecule.

6.   A recombinant vector expressing a nucleotide sequence
        encoding a pre-*trans*-splicing molecule.

5   7.   The molecule of claim 1 further comprising a safety nucleotide
                sequence comprising one or more complementary sequences
                that bind to one or more sides of the pre-*trans*-splicing
                molecule branch point, pyrimidine tract, or 3' splice site.

        8.   A method of producing a chimeric mRNA molecule comprising
10   contacting a pre-*trans*-splicing molecule with a target pre-
                mRNA under conditions in which a portion of the pre-*trans*-
                splicing molecule is *trans*-spliced to a portion of the target pre-
                mRNA to form said chimeric mRNA.

        9.   The method of claim 8 wherein the pre-*trans*-splicing mRNA
15   comprises nucleotide sequences encoding the cystic fibrosis
                trans-membrane regulator.

        10.   The method of claim 8 wherein the pre-*trans*-splicing mRNA
                comprises exon 10 cystic fibrosis trans-membrane regulator
                nucleotide sequences.

20

        11.   A method of providing a host cell with a chimeric mRNA
                molecule, said method comprising:
                transferring a pre-*trans*-splicing molecule to a host cell
                expressing a target pre-mRNA wherein the pre-*trans*-splicing
25   molecule binds to the target pre-mRNA under conditions in

53

which a portion of the pre-*trans*-splicing molecule is *trans*-spliced to a portion of the target pre-mRNA to form a chimeric mRNA molecule.

12. The method of claim 11 wherein the host cell is a human cell.

5

13. The method of claim 12 wherein the pre-*trans*-splicing molecule comprises nucleotide sequences encoding a protein that is defective or lacking in the host cell.

14. The method of claim 13 wherein the pre-*trans*-splicing molecule comprises nucleotide sequences encoding a fragment

10 of the cystic fibrosis trans-membrane regulator protein.

15.. A pharmaceutical composition comprising a pre-*trans*-splicing molecule and a pharmaceutically acceptable carrier.

FIGURE I.A

Binding Domain   Spacer   Splice Site   Delivered Therapeutic Gene

(B)   (1) pPTM+Sp

(2) pPTM-Sp

(C)

A

B

CTAG

G
A
G
A
T
G
T
T
C
C
A
G

Exon 1 of βIICG6

1st coding nucleotide of DT-A

G
G
C
G
T
G
A
T
G
A
T
G
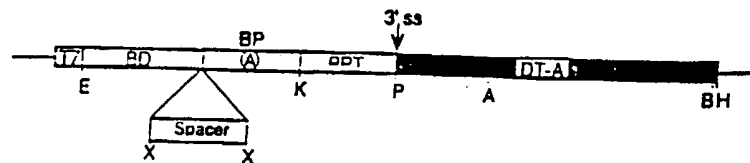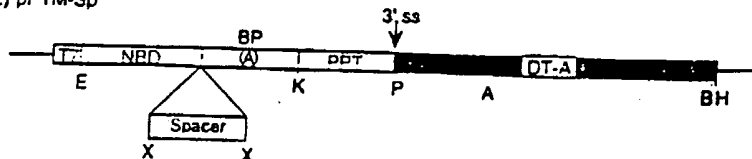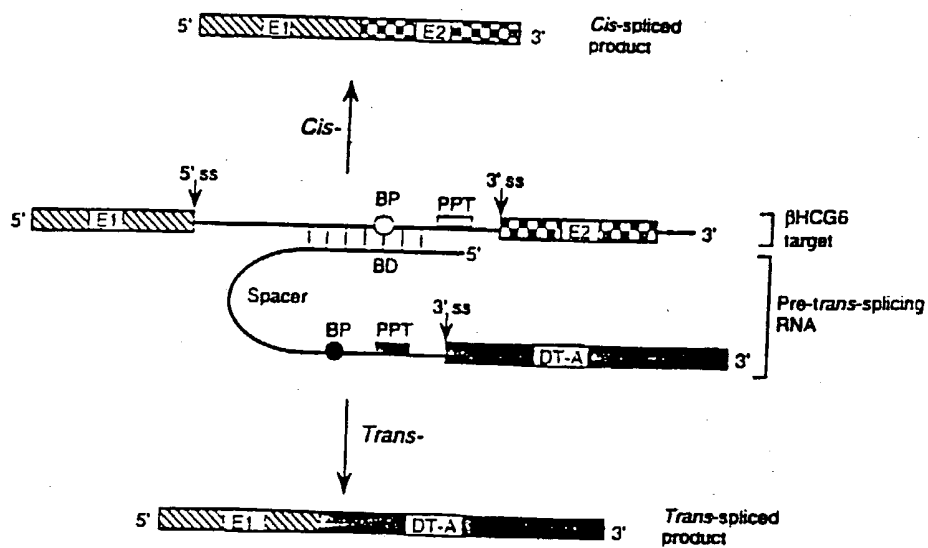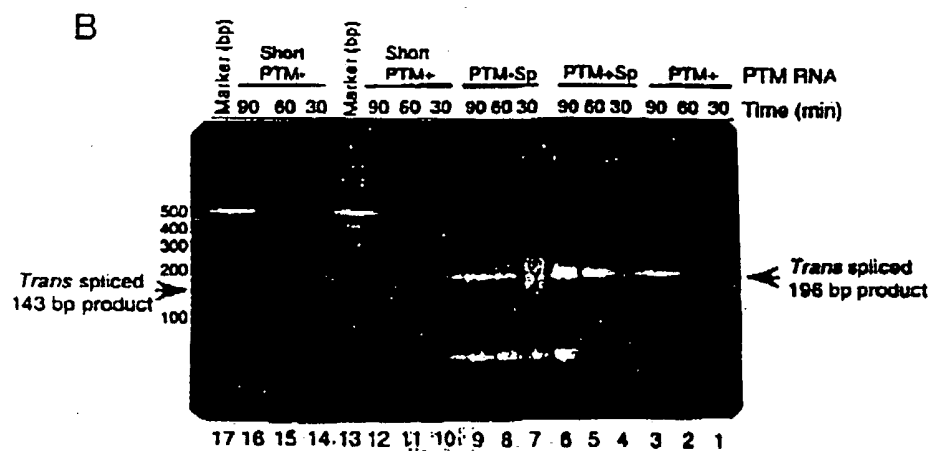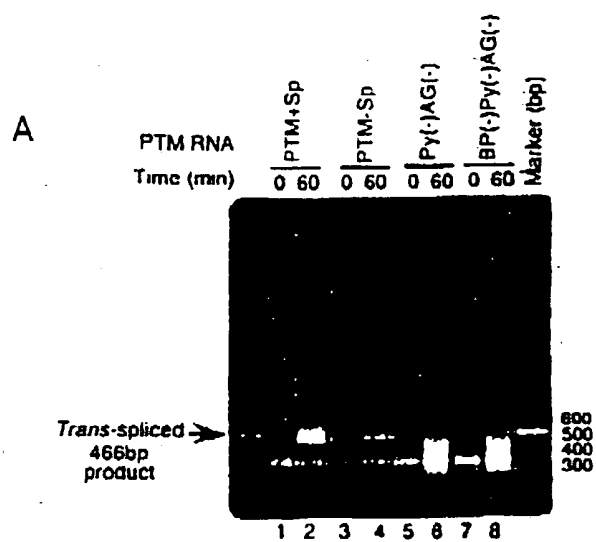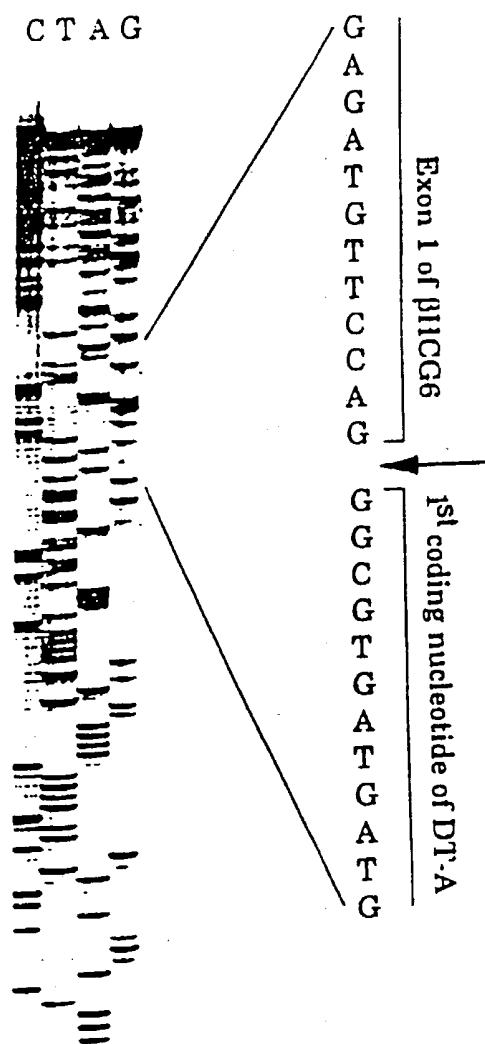
(A)
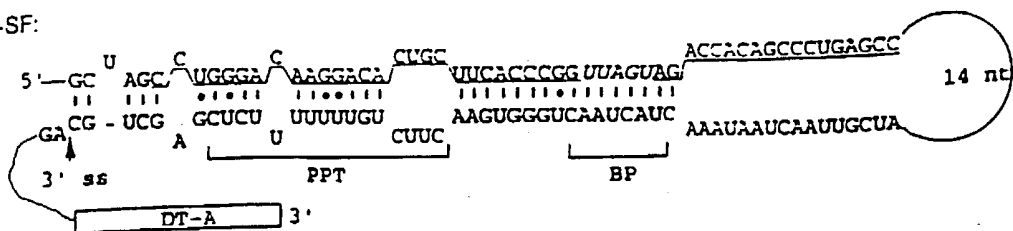
1. PTM+SF:

2. PTM+SF-Py1:

3. PTM+SF-Py2:

(B)

Figure 4

(C)

Trans-spliced
196 bp product

	Linear PTM							Safety PTM				
Marker (bp)	βHCG-F		β-globin-F			Marker (bp)	βHCG-F		β-globin-F			Forward Primer
	HCGR2	DT-3R	β-globin-R	β-globin-R	DT-3R	HCGR2		HCGR2	DT-3R	β-globin-R	β-globin-R	DT-3R	HCGR2	Reverse Primer

500
400
300
200

100

1 2 3 4 5 6   7 8 9 10 11 12

(A)

(B)

Un-spliced 478 bp

Cis-spliced 125 bp product

Trans-spliced 196 bp product
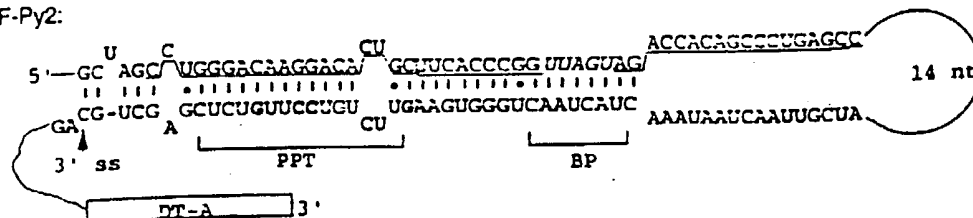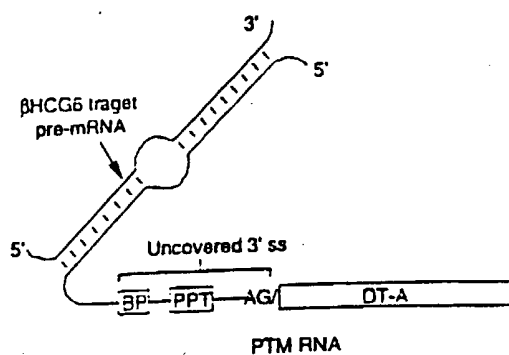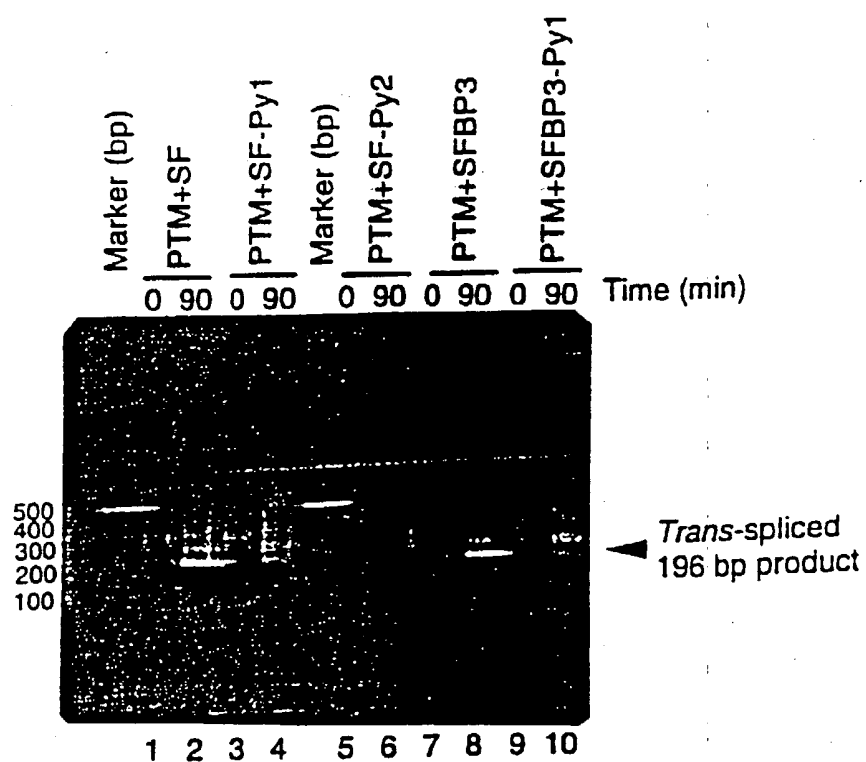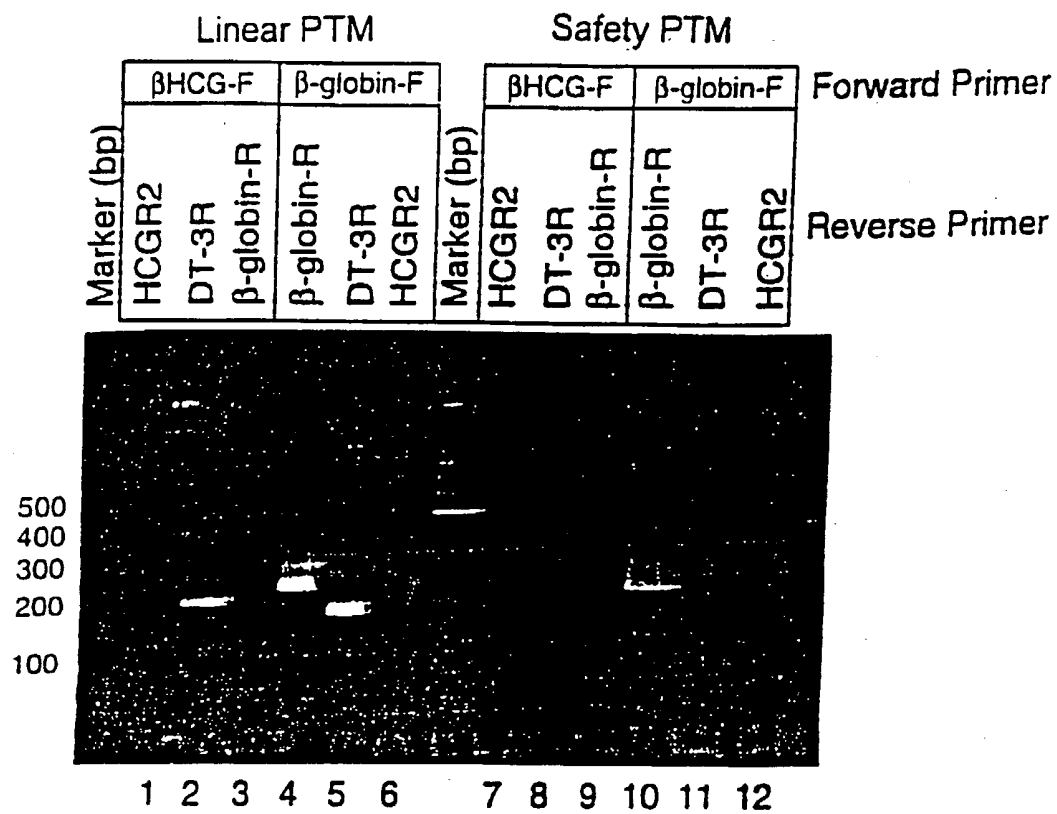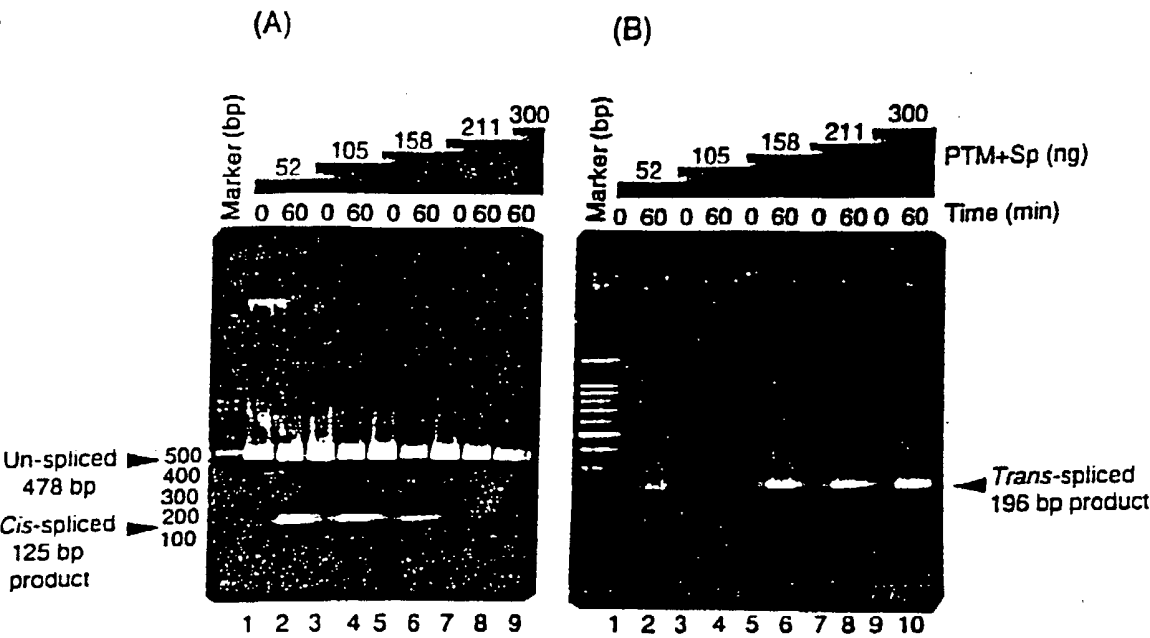
Figure 7

(A)

(B)
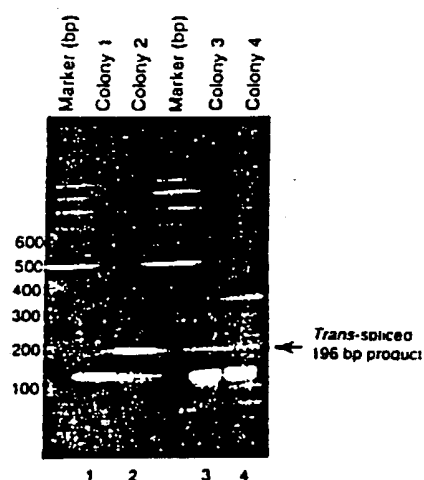
Exon 1 of βHCG6 ▼
5'-CAGGGGACGCACCAAGGATGGAGATGTTCCAG-GGCGCTGATGATGTTGTT
▲ 1st coding nucleotide of DT-A

GATTCTTCTTAAATCTTTTGTGATGGAAAACTTTTCTTCGTACCACGGGACTA

AACCTGGTTATGTAGATTCCATTCAAAAA-3'

Figure 8A

Double Splicing Pre-therapeutic RNA
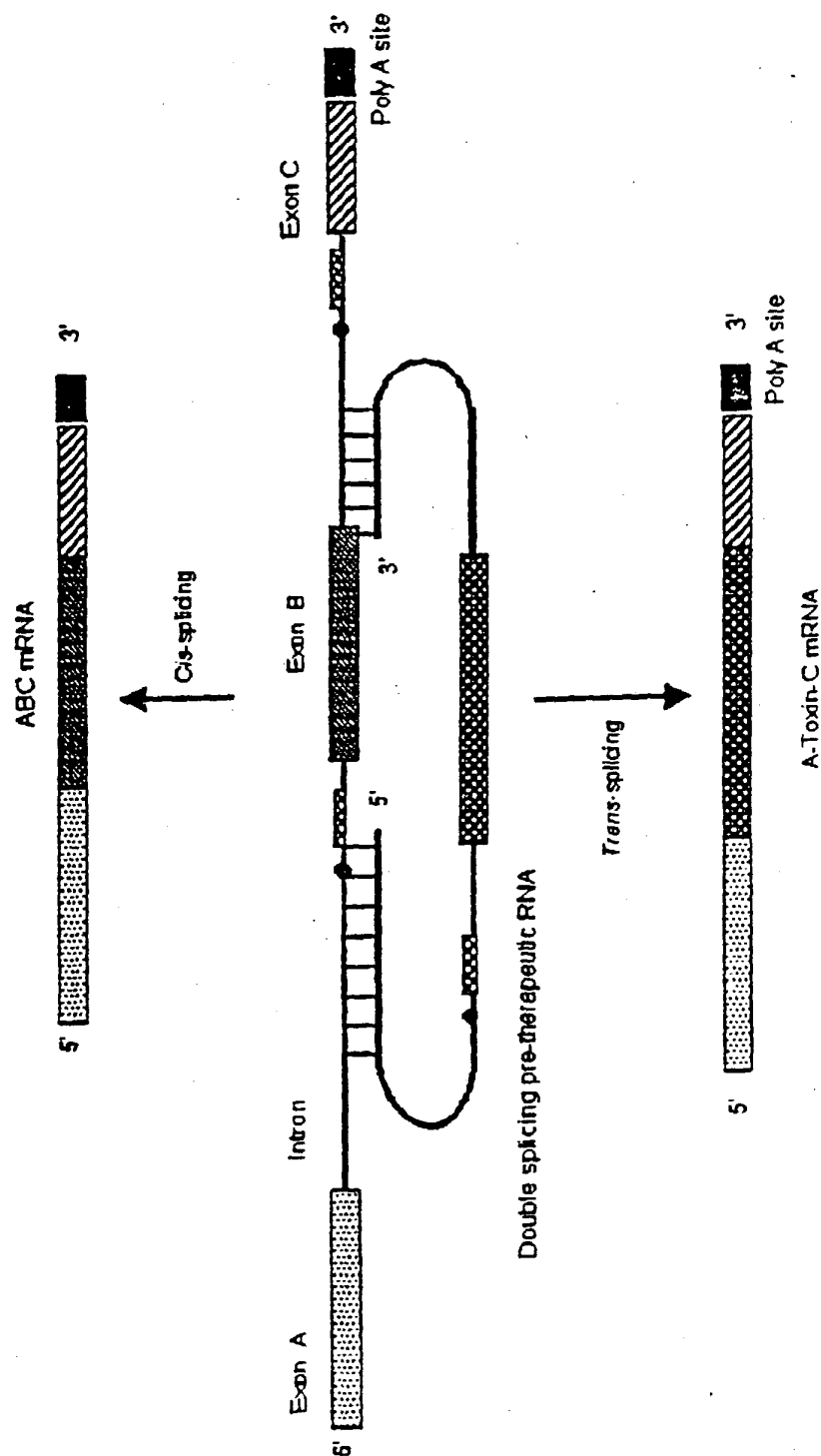
Selective _Trans_-splicing of a Double Splicing PTM

(3' ss of PTM to 5' ss target and, 5' ss of PTM to 3' ss of target)

Figure 8B
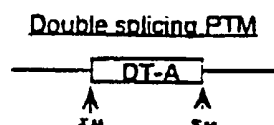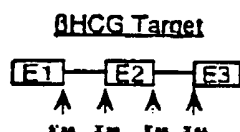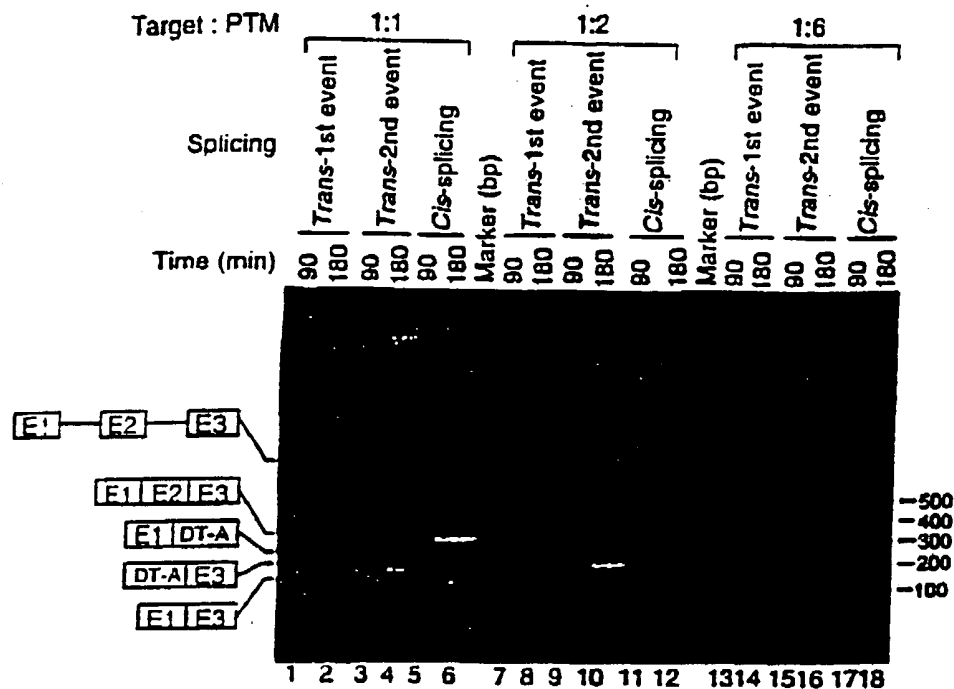
Cis-spliced products

E1 E2 E3 = Normal _cis_-splicing (277bp)

E1 E3 = Exon skipping (110bp)

Trans-splicied products

E1 DT-A = 1st event, 196bp. _Trans_-splicing between 5' ss of target & 3' ss of PTM.

DT-A E3 = 2nd event, 161bp. _Trans_-splicing between 3' ss of target & 5' ss of PTM.

genomic DNA

pre-mRNA

intron   exon   intron

trans-splicing

PTMs

1. RT with random oligos
2. PCR with specific oligo to
   make one strand
3. Second strand synthesis
   (questions here)

N

4. NotI cleavage and cloning

5. Random short sequence
   runs

Mapping of 3' splice sites

1. RT with specific oligos
   second strand synthesis

N

2. NotI cleavage and cloning

3. Random short sequence
   runs

Mapping of 5' splice sites
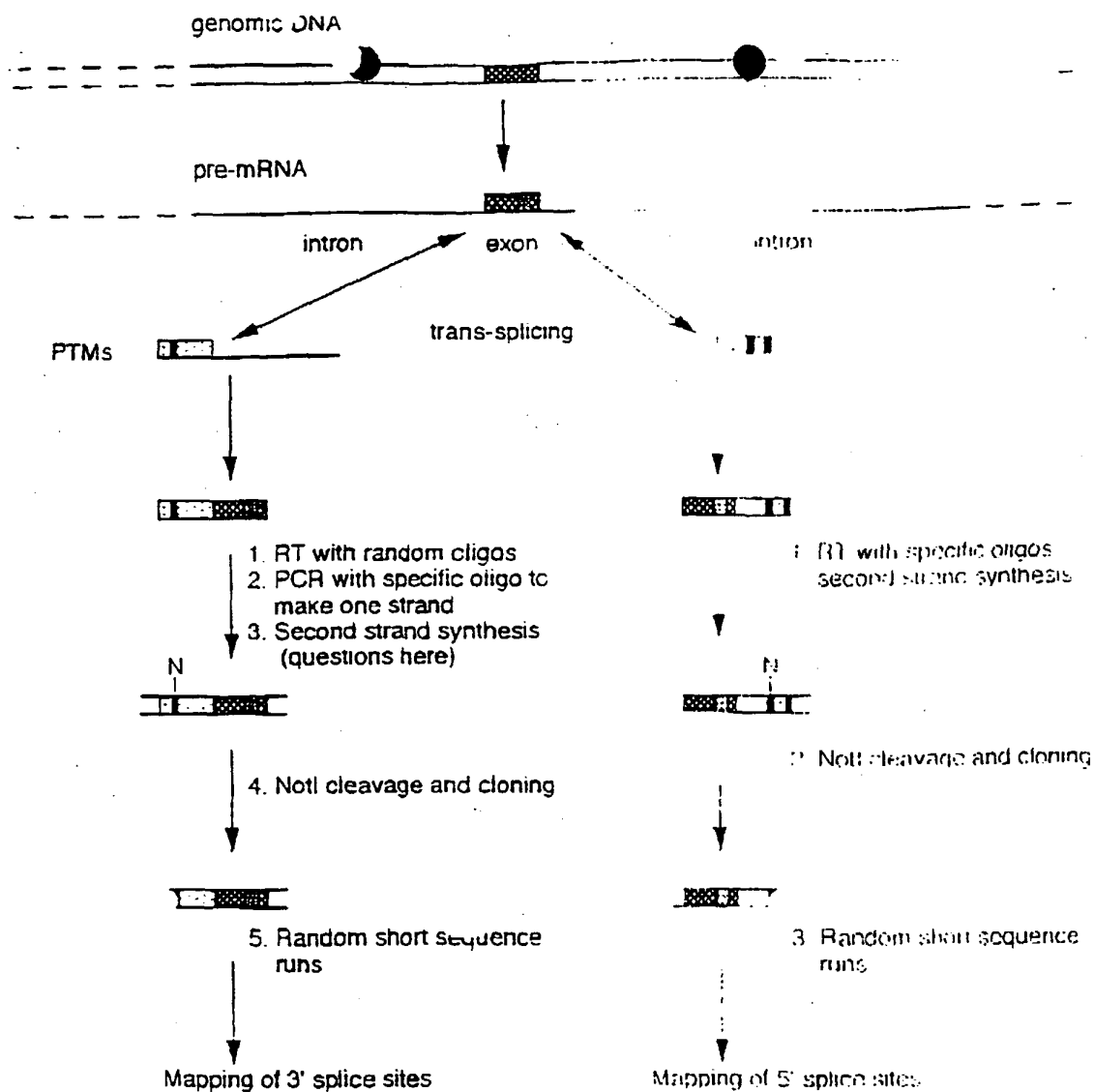
FIGURE 9

FIG. 10 A

Knock Out
LacZ ~~Inform~~ Model Constructs

pc3·1lac-TI
Target 1:

PTMs

pc3·1PTM2 :-
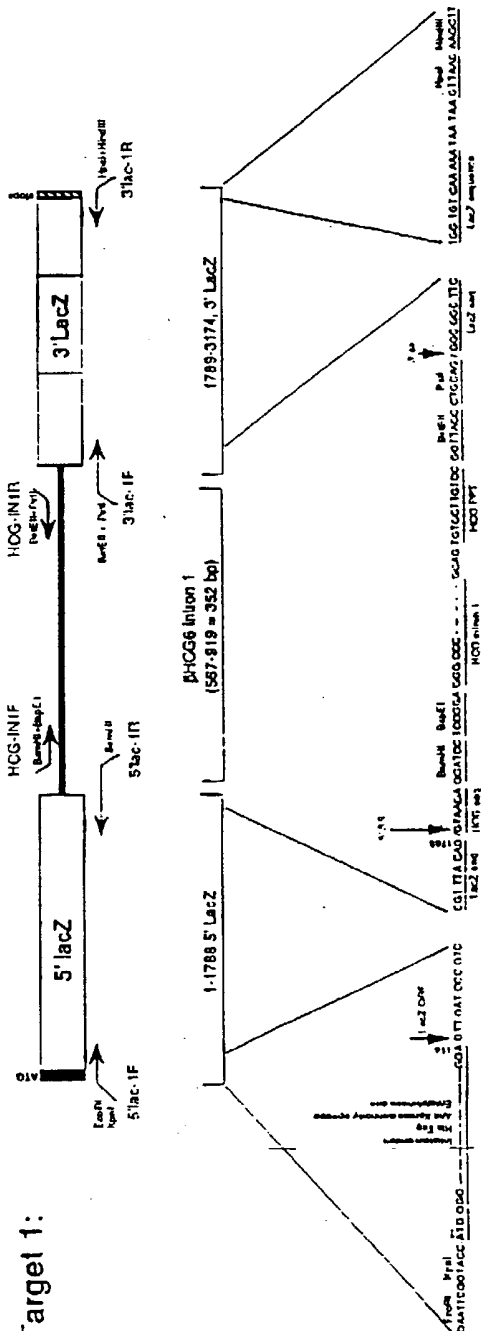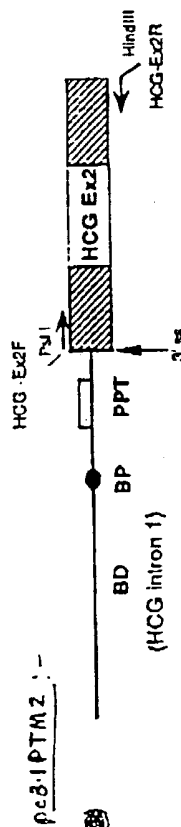
Restoration of β-Gal activity by SMaRT
(Spliceosome Mediated RNA Trans-splicing)
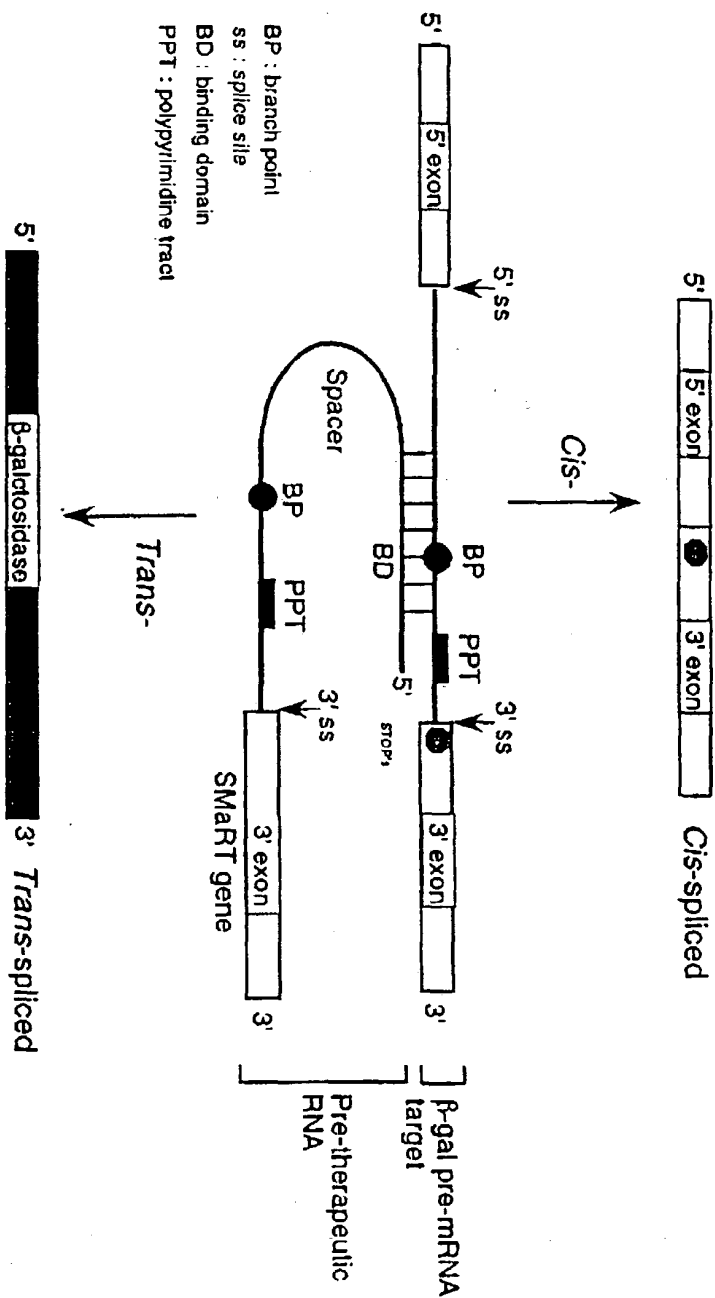
Figure 10 B
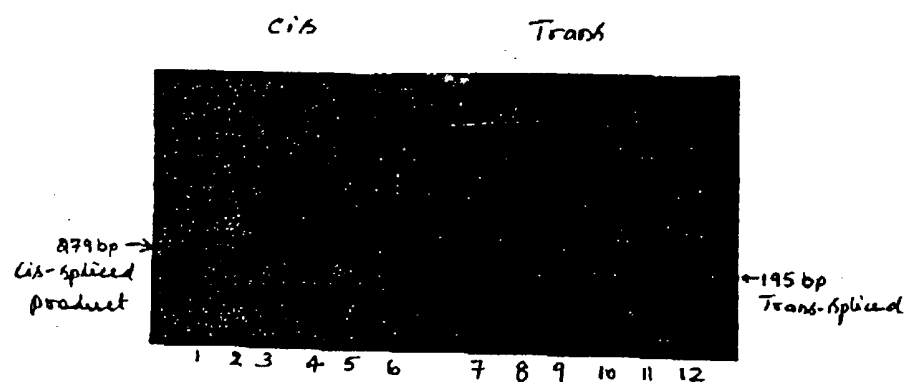
FIGURE 11A

Figure 11 B

FIGURE 11C

Nucleotide Sequence Demonstrating that
Trans-splicing is Accurate

G A T C

```
                                    ┌ C ┐
                                    │ A │
                                    │ T │
                                    │ A │
                                    │ G │
                                    │ G │
                                    │ G │
                                    │ G │
                                    │ C │ LacZ 5' fragment
                                    │ A │
                                    │ A │
                                    │ A │
                                    │ T │
                                    │ G │
                                    │ T │
Trans-splice                        │ C │ ◄── Trans-splice
junction  ──────►                   │ C │     junction
                                    │ C │
                                    │ C │
                                    │ G │
                                    │ A │ HCG exon 2
                                    │ C │
                                    │ G │
                                    │ A │
                                    └ C ┘
```
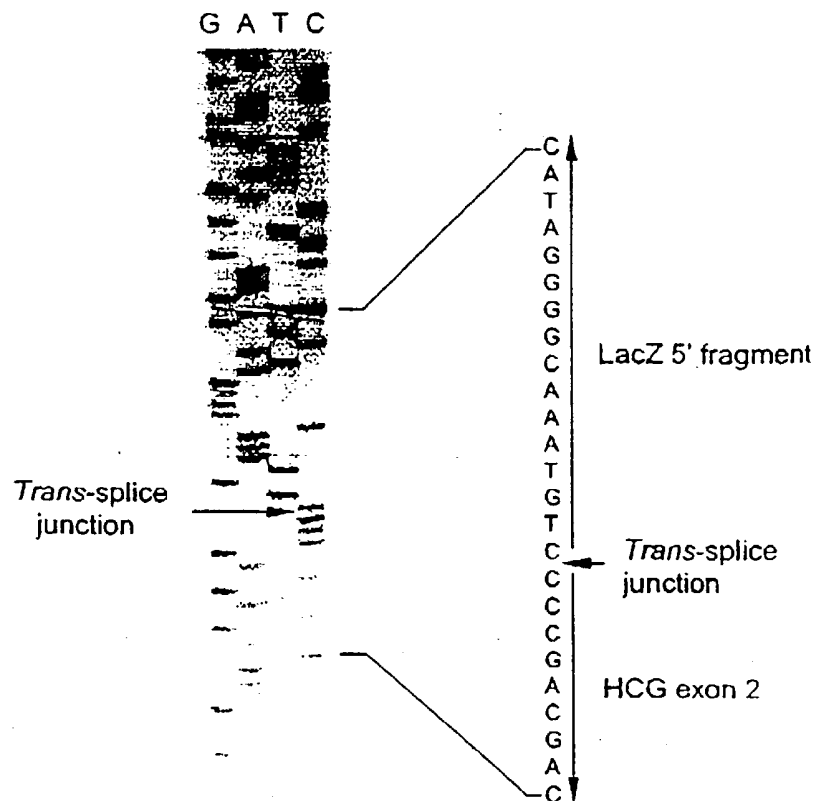
FIGURE 12 A

(1). Nucleotide sequences of the *cis*-spliced product (285 bp) :

BioLac-TR1
GGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCCACGCGATGGGTAACAGTCTTG

                                                        Splice junction
GCGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAG/GGCGGCTTCGTCTAATAATG

GGACTGGGTGGATCAGTCGCTGATTAAATATGATGAAAACGGCAACCCGTGGTCGGCTTACGGCGGTGATTT

                                                                Lac-TR2
TGGCGATACGCCGAACGATCGCCAGTTCTGTATGAACGGTCTGGTCTTTGCCGACCGCACGCCGCATCCAG

(2) Nucleotide sequences of the trans-spliced product (195 bp)

BioLac-TR1
GGCTTTCGCTACCTGGAGAGACGCGCCCGCTGATCCTTTGCGAATACGCCCACGCGATGGGTAACAGTCTTGG

                                                        Splice junction
CGGTTTCGCTAAATACTGGCAGGCGTTTCGTCAGTATCCCCGTTTACAG/GGGCTGCTGCTGTTGCTGCTGCT

                                                HCGR2
GAGCATGGGCGGGACATGGGCATCCAAGGAGCCACTTCGGCCACGGTGCCG

Figure 12 B

CFTR Pre-therapeutic molecule (PTM or "bullet")
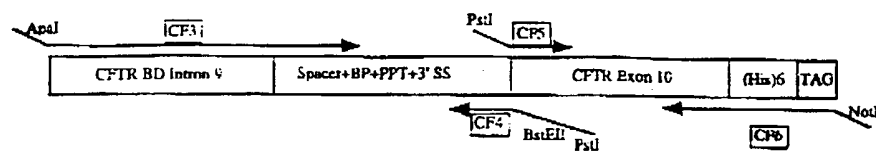
CFTR mini-gene target - Construction
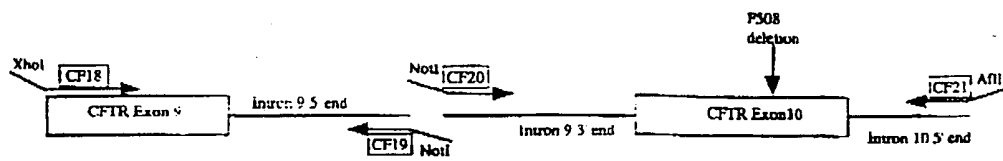
TRANS- SPLICING Repair
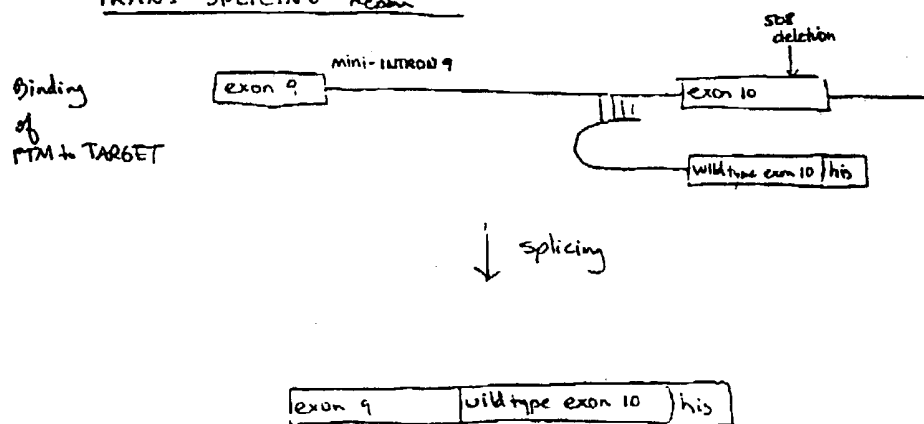
Figure 13

Figure 14
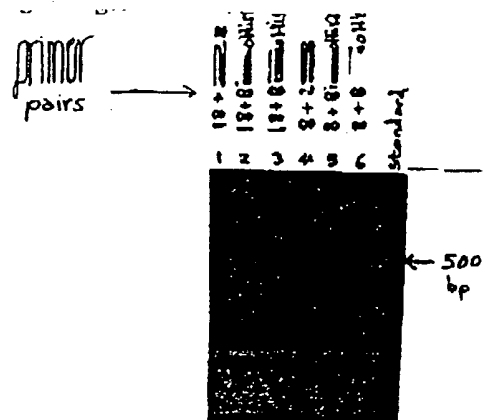
DNA sequence   500 b.p.   GCTAGCGTTTAA ... TGCCACTCCCAC   linear

Positions of Restriction Endonucleases sites (unique sites underlined)

```
                Sau96 I
                Hae III
                Sau96 I
                Ban II                    Binding domain
Nhe I  Dra I   Apa I     Intron 9  BD          Sac II                    Xho I   Sca I
|      |       ||     |                        |                        |       |
GCTAGCGTTTAAACGGGCCCACCCATCATTATTAGGTCATTATCCGCGGAACATTATTATAACGTTGCTCGAGTACTAAC
CGATCGCAAATTTGCCCGGGTGGGTAGTAATAATCCAGTAATAGGCGCCTTGTAATAATATTGCAACGAGCTCATGATTG   80
|      |       ||     |                        |                        |   |
1      8       15                              44                       68  |
               15                                                           72
               15
               16
               16
```

```
Kpn I                    Pst I          Exon 10  CFTR + His tag + Stop
|                        | 3'ss
TGGTACCTCTTCTTTTTTTTTCCTGCAGACTTCACTTCTAATGATGATTATGGGAGAACTGGAGCCTTCAGAGGGTAAAAT
ACCATGGAGAAGAAAAAAAAAGGACGTCTGAAGTGAAGATTACTACTAATACCCTCTTGACCTCGGAAGTCTCCCATTTTA   160
|                        |
82                       102
```

```
            Xmn I              Dde I                                          F508
            |                  |                                             |
TAAGCACAGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGGATTATGCCTGGCACCATTAAAGAAAATATCATCTTTG
ATTCGTGTCACCTTCTTAAAGTAAGACAAGAGTCAAAAGGACCTAATACGGACCGTGGTAATTTCTTTTATAGTAGAAAC   240
            |                  |
            172                190
```

```
                                    Sph I                         His        STop
                                    |                                        ↓
GTGTTTCCTATGATGAATATAGATACAGAAGCGTCATCAAAGCATGCCAACTAGAAGAGCATCATCATCATCATCATTAG
CACAAAGGATACTACTTATATCTATGTCTTCGCAGTAGTTTCGTACGGTTGATCTTCTCGTAGTAGTAGTAGTAGTAATC   320
                                    |
                                    282
```

```
                                             Sac I
                                             Ban II
                                             Sau3A I
            Hae III            Pst I         Dpn I
            Not I              EcoR V  EcoR I BamH I         Kpn I    HinD III
            |                  |       |     ||             |        |        Dra I
GCGGCCGCCACTGTGCTGGATATCTGCAGAATTCCACCACACTGACTACTGCGATCCGAGCTCGGTACCAAGCTTAAGTT
CGCCGGCGGTGACACGACCTATAGACGTCTTAAGGTGGTGTGACTGATCACCTAGGCTCGAGCCATGGTTCGAATTCAA   400
||                 |       |           |          |             |        |
321                339     349         | CF28 372  |             384      399
323                        344                373                390
                                                 373
                                                       378           Present in PTM 3'uT
                                                       378           but not Target.
```

```
Sau3A I
Dpn I
|
TAAACCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCGTGCCTTCCTTGACC
ATTTGGCGACTAGTCGGAGCTGACACGGAAGATCAACGGTCGGTAGACAACAAACGGGGAGGGGCACGGAAGGAACTGG   480
|
410                    CF27
410
```

```
CTGGAAGGTGCCACTCCCAC    500
GACCTTCCACGGTGAGGGTG
```

Restriction Endonucleases site usage

Acc I	-	EcoR I	1	Nda I	-	Sau96 I	2
Apa I	1	EcoR V	1	Nhe I	1	Sca I	1
ApaL I	-	Hae II	-	Not I	1	Sma I	-
Avr II	-	Hae III	2	PflM I	-	Sph I	1
BamH I	1	HinC II	-	Pst I	2	Spl I	-
Ban II	2	HinD III	1	Pvu I	-	Ssp I	-
Bbe I	-	Hinf I	-	Pvu II	-	Stu I	-

EXPERIMENT 12

Repair of an exogenously supplied CFTR target
molecule carrying an F508 deletion in exon 10.

PTM | CFTR 5' intron 9 | Spacer+BP+PPT+3' SS | CFTR exons 10-24 | (His)6 | TAG |

+

CFTR Target
(mini-gene)

F508 deletion

| CFTR Exons 1-9 | CFTR Exons 10-24 |

Mini-intron 9
(~0.6 kb)

Cotransfect PTM and Target molecules in HEK 293 cells
and detect repaired CFTR mRNA by RT-PCR.

Repaired
CFTR mRNA

| CFTR Exons 1-9 | Exon 10-24 CFTR | (His)6 | TAG |

EXPERIMENT 3

Repair of endogenous CFTR
transcripts by exon 10 invasion
using a double splicing PTM

Double Splicing
PTM

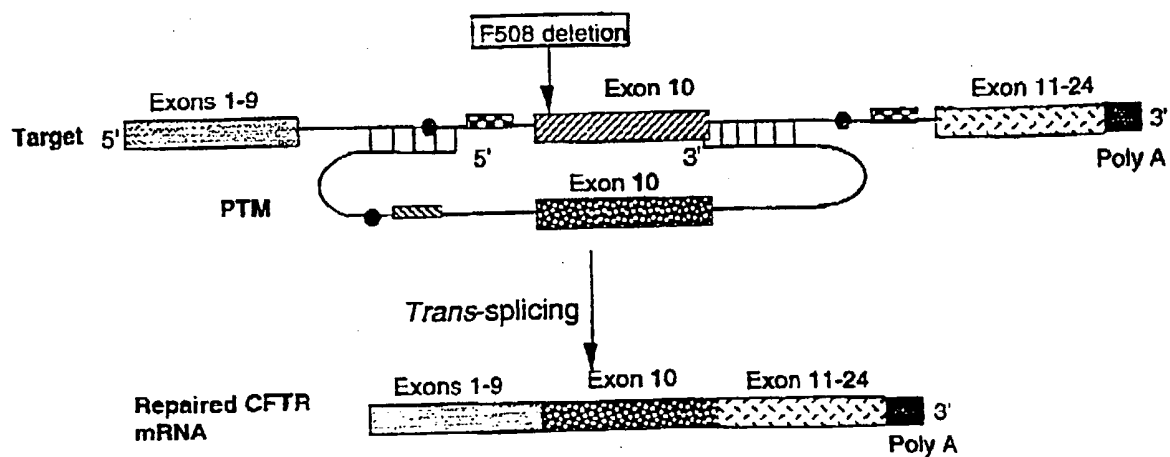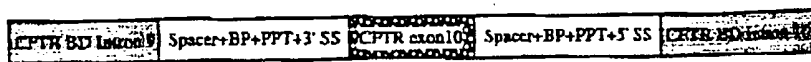| CFTR BD Intron 9 | Spacer+BP+PPT+3' SS | CFTR exon10 | Spacer+BP+PPT+5' SS | CFTR BD Intron 10 |

F508 deletion

Exons 1-9   Exon 10   Exon 11-24
Target 5'   3'
                                                    Poly A
         PTM   5'   Exon 10   3'
                             Exon 10

Trans-splicing

Repaired CFTR   Exons 1-9   Exon 10   Exon 11-24
mRNA   3'
                                              Poly A

Figure 17